

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

## BEE LINE

Beeswax traces in pottery vessels date human use of hive products to the beginnings of agriculture **PAGE 226**

### SOCIAL MEDIA

#### BREAKING BOUNDARIES

Twitter takes on sexism in science

**PAGE 148**

### BIOHAZARDS

#### BIOSAFETY CULTURE

What life scientists can learn from the nuclear industry

**PAGE 155**

### ASTROPHYSICS

#### A PLANET TO WATCH

A nearby Earth-size body with atmosphere

**PAGES 169 & 204**

**NATURE.COM/NATURE**

12 November 2015 £10

Vol. 527, No. 7577





# THIS WEEK

## EDITORIALS

**FUNDING** Arecibo's woes are emblematic of tough times **p.134**

**WORLD VIEW** A plan to shut down China's legal ivory trade **p.135**

**MATERIALS** Flexible battery made from paper, ink and foil **p.137**



## A breath of fresh air

*The decision to use the Montreal Protocol to reduce the impact of refrigerants on global warming is a step forward ahead of the Paris climate summit.*

The world took a step in the right direction in the early hours of 6 November. Meeting in Dubai, 195 governments decided to pull the management of hydrofluorocarbons (HFCs) under the umbrella of the Montreal Protocol, an international agreement that governments signed in 1987 to protect the stratospheric ozone layer from damage by chlorofluorocarbons (CFCs). HFCs are ozone-friendly replacements for CFCs and are often used as refrigerants. Unfortunately, many are also powerful greenhouse gases, and their use is expected to skyrocket over the coming decades. Although a few governments were successful in delaying negotiation of the arrangement's details, the agreement to move forward is nonetheless welcome — and long overdue.

The United Nations Environment Programme estimates that HFC emissions are rising roughly 7% annually thanks in part to demand for air conditioners in emerging economies such as Brazil, India and Indonesia. By 2050, global HFC emissions could hit the equivalent of 8.8 billion tonnes of carbon dioxide, roughly equal to the current carbon emissions of the United States and European Union combined. The decision to regulate HFCs under the Montreal Protocol bodes well for the UN climate summit in Paris, which begins on 30 November. The Montreal Protocol is a well-oiled machine that has already proved its effectiveness and value among both industry and government leaders. In theory, dragging HFCs under its umbrella gives climate negotiators in Paris one less thing to worry about. All told, aggressively regulating HFCs could

reduce global warming by an estimated 0.5 °C.

Unfortunately, some governments still seem to be hedging their bets. In particular, India, Saudi Arabia and Kuwait pushed to delay consideration of detailed amendments to the Montreal Protocol until next year. Other global leaders, including US President Barack Obama, who has negotiated agreements on HFCs with leaders in India, China and Pakistan, will need to maintain pressure.

A day after his administration helped to negotiate the agreement on the Montreal Protocol, Obama denied permission for construction of the Keystone XL pipeline from Canada to the southern United States (see page 141). His decision was based on a simple criterion: whether the project would be in his country's 'national interest'. Obama said the answer is no, and he may be right. The world is awash with oil at the moment, and there is no need to rush any of it to market, let alone carbon-rich crude from Alberta's tar sands.

The world must, however, be clear about what this decision does — and does not — accomplish. It sends a signal to industry about the environmental values of the current administration, and it may make things a little harder for Canadian companies seeking to develop and export dirty crude oil. But a symbolic one-off gesture from one government will not change the fundamental dynamic driving greenhouse-gas emissions. To do that, all countries must implement meaningful policies that will rush low-carbon technologies to market. The clean-energy pipeline runs through Paris and next year's Montreal Protocol meetings on HFCs. ■

## Universities' value

*Proposals for UK higher education contain some positive points amid the financial gloom.*

The UK government's austerity policies are soon expected to deliver swingeing cuts in some departments. In the teeth of those prospects, British researchers, and an influential parliamentary science and technology committee, have lobbied hard to make the case that even a flat research budget, after five continuous years of the same, would be a betrayal of the country's needs.

How successful they have been, and how worthy of exception the government considers them to be, will become clear only when UK spending plans are announced on 25 November. On page 144, we explore the worrisome prospects.

As if that wasn't enough, last week the government began a consultation over its proposed restructuring of the way it administers higher-education funding (see [go.nature.com/c97sww](http://go.nature.com/c97sww)). It

announced that it wants to abolish the Higher Education Funding Council for England (HEFCE), the body that distributes £1.6 billion (US\$2.4 billion) of 'quality related' research money to universities. Those core funds would still be administered separately from the more-responsive funding by the UK research councils, probably by a research-funding organization that would be responsible for both areas.

**"The impacts case studies provide welcome ammunition to the case for supporting research in all disciplines."**

The case for those particular changes has not been adequately made, but other aspects of the proposals have virtues. The government is tackling two scandals in the UK higher-education system: its relative neglect of quality standards in teaching, and its inadequacies in contributing to social mobility.

Another positive feature is that the government supports the continuation of the Research Excellence Framework (REF) for assessing research quality and impact, despite the proposed abolition of HEFCE, which successfully implemented that process. It is worth taking stock of the REF — not least because its results strengthened the case for



research investment by government.

True, many academics hated the REF, which required them to submit large quantities of information to justify their funding. But an inspection of the REF's outcomes, and of the retrospective reviews of the process by international members of the REF assessment panels (see [go.nature.com/q919oe](http://go.nature.com/q919oe)), suggests that it has many strengths.

Take the database of nearly 7,000 case studies of the societal impacts of academic research (see <http://impact.ref.ac.uk/casestudies>). The diversity of the impacts in terms of (for example) health, sustainability, education and economic growth — in the United Kingdom and beyond — is remarkable and inspiring.

And there is no reason to suppose that identifying these outcomes is the equivalent of the impacts tail wagging the research dog. In the REF, the assessment panels gave the societal case studies a mere 20% weighting, whereas academic performance had a 65% weighting. No one could sensibly maintain that the outcomes are necessarily predictable and should be required as a basis for funding in future. What does make sense is that the research community can help to ensure a maximal return on taxpayers' money by becoming aware of impacts pathways, and by broadening its outlook on its roles. As the case studies show, this can happen even in areas of research that

are unapologetically fundamental.

A study by the independent consultants RAND Europe estimates that the REF impacts-submissions process cost universities about £55 million (see [go.nature.com/dzwbjn](http://go.nature.com/dzwbjn)). That may seem to justify the concerns of academics and politicians about the burden. But set against the £1.6-billion budget that it relates to, one might even describe this 3.4% overhead as a bargain — especially given that the assessment system may become more efficient, and given the virtues of encouraging such impacts.

Mindful of the burdens, the government is evidently tempted to try to find a cheaper method of assessment of both academic and societal impacts using metrics. As is made clear by the REF panels, by a RAND analysis of the impacts evaluation ([go.nature.com/yysa6m](http://go.nature.com/yysa6m)) and by an independent assessment of metrics in research ([go.nature.com/rfrgql](http://go.nature.com/rfrgql)), this temptation should be avoided. Insightful review of both types of outputs is the only way to do justice to them.

The impacts case studies provide welcome ammunition to the case for supporting research in all disciplines, and some government departments have been deploying them in that spirit. Readers who care about UK higher education should give their own responses to the proposals before 15 January 2016, at [go.nature.com/l3rrtx](http://go.nature.com/l3rrtx). ■

# Radio interference

*Conflict at the Arecibo Observatory highlights the need for funders to become more flexible.*

In more than 50 years of operation, the Arecibo Observatory has enabled some momentous discoveries. Researchers at the observatory in Puerto Rico found in 1965 that the length of a day on Mercury is more like two Earth months than three. The first binary pulsar, detected at Arecibo in 1974, earned its discoverers the 1993 Nobel Prize in Physics. The observatory's telescope made the first confirmed discovery of extrasolar planets in the early 1990s. And two years ago, astronomers reported using Arecibo to find Leo P, a dwarf galaxy that had gone undetected just 2 megaparsecs from the Milky Way.

Unfortunately, that five-decade streak of discoveries could be near its end. Arecibo is long in the tooth and certainly shows its age, but the problem is not that the observatory has exhausted its scientific potential. Its telescope remains the largest single-dish radio instrument in the world, and is in demand from those who study topics such as pulsars, asteroids and Earth's upper atmosphere.

The problem is money, as it so often is. Operating Arecibo costs about US\$12 million per year, a price that its owner — the US National Science Foundation (NSF) — feels increasingly unable to pay.

Arecibo's bell began to toll almost a decade ago, when a panel charged with balancing the NSF's astronomy budget put the observatory on its list of potential cuts. The agency's astronomy division has long been under intense pressure to support the development of new instruments such as the Large Synoptic Survey Telescope, currently under construction in Chile, without any increase in budget.

But few in the astronomy community or the NSF seem to relish the idea of closing a facility as scientifically productive as Arecibo. So the agency has endeavoured to find parties willing to contribute to the observatory's operation or, better yet, to take over the enterprise entirely.

In July, a glimmer of hope appeared in the form of the Breakthrough Listen project, a \$100-million effort sponsored by Internet billionaire Yuri Milner to scan the nearest 1 million stars for signals suggesting the presence of intelligent life. The project has already arranged to purchase about 20% of the observing time at another NSF radio facility, the Green Bank Telescope in West Virginia.

But no similar agreement has been reached at Arecibo. One

stumbling block may have been conflicts with the observatory's director, Robert Kerr, who has said that the NSF told him that it would cut its contribution to Arecibo by an amount equal to any contribution from Breakthrough Listen. Such an arrangement, he feared, would give the agency an opportunity to absolve itself entirely of the need to fund the telescope (see page 142).

NSF officials vigorously deny any such intent, and insist that they never even suggested a one-for-one reduction of funding as a condition of a Breakthrough Listen deal. Nevertheless, Kerr's objections led to his resignation as operations director. On 26 October, the science agency distributed an open letter soliciting "strategies and goals for continued operations that involve a substantially reduced funding commitment from NSF".

It might seem difficult to understand why the NSF has so much trouble accepting help from a private source. But the agency deserves some sympathy. It has a mandate to support a research agenda dictated by the scientific community through peer review, not by the interests of deep-pocketed philanthropists.

That is not to say that Breakthrough Listen's goals lack legitimacy — the argument that intelligent life must exist somewhere else in the universe is strong, the confirmation of a signal from an alien civilization would arguably be the greatest scientific discovery ever made and the Arecibo telescope is one of the world's best instruments for making such a discovery. But accepting Breakthrough Listen's offer without minimizing its impact on the general community's access to the telescope would be irresponsible. The NSF is right to proceed slowly.

Such episodes are likely to be repeated as long as public science funding remains tight and the technology sector continues minting billionaires with the curiosity and intellect to put their money to work expanding the frontiers of human knowledge. In the case of Arecibo, and in future disagreements over how to best combine private and public resources, it is important to remember that although they may differ on how exactly to proceed, all sides share an interest

in the advancement of science. Elected officials are unlikely to increase funding for research any time soon, so it would be wise for managers at funding agencies and at publicly funded institutions to embrace the spirit of compromise. ■

***"Although they may differ on how exactly to proceed, all sides share an interest in the advancement of science."***

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqv](http://go.nature.com/xhunqv)





## China must act decisively to eradicate the ivory trade

*The continued existence of large stockpiles of legal raw ivory in the country is hampering China's promises to save the elephant, says Li Zhang.*

**T**he fight against the illegal ivory trade had some success last month with the arrest of a Chinese woman alleged to have smuggled more than 700 elephant tusks from Tanzania. Ivory is a particular prize in China, where it has been carved into ornaments, chopsticks and jewellery for more than 3,000 years.

This event, and many similar ones, puts the Chinese government in an awkward position. Officially, China has pledged to help to prevent the illegal killing of elephants, and has taken steps to stop the import of illegal ivory. Yet in 2006, the Chinese Ministry of Culture designated ivory carving a national intangible cultural heritage. As such, the government has continued to support the domestic sale and use of 'legal' ivory — and occasionally buys global stockpiles of confiscated ivory to sell on to its carvers. There are 34 officially designated ivory manufacturers in China and 130 licensed dealers who sell tonnes of legal ivory items to shoppers each year.

The Chinese government has argued that there are enough elephants in Africa to support regular — and legal — sustainable trade to meet demand, and that its domestic market therefore does not lead to increased poaching. It introduced regulations requiring each ivory product to be registered and sold with a certificate showing that it is from an approved source. Buoyed by official support, the ivory industry promotes its products to the public as a wise financial investment. In 2009, after a controversial approval process by CITES, the Convention on International Trade in Endangered Species of Wild Fauna and Flora, China bought and imported 62 tonnes of ivory from countries across southern Africa.

The problem with this approach is that the legal market encourages poaching and the continued introduction of illegal ivory into the market. Surveys in China suggest that one-third of the licensed stores still sell illegal ivory, and that dozens of unlicensed dealers continue to operate.

Recorded animal deaths and seizures of illegal ivory confirm the scale of the continuing black market. Tens of thousands of elephants are poached each year, and much of the resulting ivory is smuggled to China and southeast Asia. More than half a tonne of seized ivory was crushed in Beijing in May, and more than six tonnes were destroyed in Guangdong in January last year.

When the ivory was crushed in May, the Chinese government said that it wanted to do more to protect the elephants, and pledged to phase out its domestic — legal — ivory trade. Last month, on a visit to the United States, China's President Xi Jinping went further and agreed a deal to take "significant and timely" steps to close down this domestic market. Popular opinion in China seems to support such a total ban on ivory sales, but the large

stockpiles — Chinese firms are reckoned to be sitting on 40 tonnes of legal raw ivory and unfinished products since the 2009 deal — must be dealt with somehow.

Chinese law, which aims to protect the newly established market economy, does not allow registered manufacturers and suppliers to be simply closed down. Instead, ivory must be decommercialized.

The simplest way to do this, and one that the government should consider, is to buy back all legally held commercial ivory stock in a compulsory-purchase scheme. The government should then pass these stockpiles to museums as appropriate. After such a move, no trade in ivory would be allowed; any ivory product on the market would be illegal. This would make enforcement simpler, cheaper and more effective. Ivory sculptures could still be displayed in museums

as national treasures, and used in schools to educate children about Chinese heritage and also about the need for conservation.

How much would such a step cost? In 2009, the purchase price for the legal ivory stock was US\$1,350 per kilogram. Given the cost of storage and management since then, a fair price now would be about \$2,100 per kilogram. As such, it would cost about \$84 million to buy back the raw ivory. Current market prices would add about another \$500 million for purchasing the worked products.

The overall cost looks high, but China regularly spends billions of dollars on what it calls ecological compensation. Farmers and landowners are paid to convert marginal cropland to forest or

grassland (the 'grain for green' scheme), and local governments are given funds for conservation and ecosystem management. Some estimates suggest that China has spent \$100 billion on such projects, so the \$600 million or so needed to eliminate the demand for ivory represents a small percentage of the overall effort. There are ways to reduce the cost if necessary. Some of the money could be used to retrain the state-owned ivory-carving manufacturers to produce alternative goods.

Under the terms of last month's presidential agreement, the United States (the second-biggest market for illegal ivory after China) said that it would tighten its own restrictions on the import of elephant products, including hunting trophies. The cultural significance and value of ivory go back centuries. As Joseph Conrad wrote in *Heart of Darkness*: "The word 'ivory' rang in the air, was whispered, was sighed. You would think they were praying to it." It will not be easy to separate this cultural value from a monetary value. But to save the elephant, we must try. ■

**Li Zhang** is a professor in the Key Laboratory for Biodiversity Science and Ecological Engineering, Ministry of Education; College of Life Sciences, Beijing Normal University.  
e-mail: asterzhang@bnu.edu.cn

THE CULTURAL  
SIGNIFICANCE  
AND VALUE  
OF IVORY  
GO BACK  
CENTURIES.

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/m2mk1p](http://go.nature.com/m2mk1p)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## NEUROSCIENCE

### Alzheimer's drugs make brains buzz

Several candidate drugs for Alzheimer's disease have failed to improve cognition in human clinical trials. One reason could be that they worsen neuronal defects, at least in mice.

The drugs are antibodies that have been designed to bind to and reduce levels of amyloid- $\beta$  protein, which builds up in the brains of people with Alzheimer's. Marc Aurel Busche and Arthur Konnerth at the Technical University of Munich in Germany and their colleagues used high-resolution imaging of mouse brains to monitor how the antibodies affect neuronal activity. In two mouse models of Alzheimer's, two different antibodies increased the number of cells in the cortex that were electrically hyperactive, further impairing brain function, compared with untreated animals.

The results suggest a greater need to test how molecular therapies affect neuronal function in the brains of living animals, the authors say.

*Nature Neurosci.* <http://dx.doi.org/10.1038/nn.4163> (2015)

## MATERIALS

### Boron made into 2D sheet

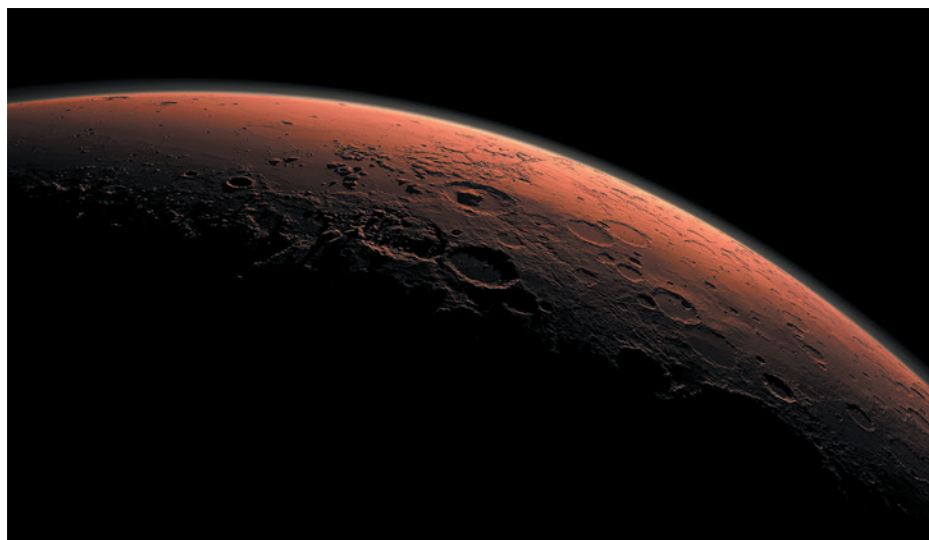
Researchers have created a 2D form of boron and shown that it behaves as a semiconductor.

Materials such as graphene (the atomically thin form of carbon) and others made of silicon and phosphorus have desirable electrical properties that could be useful in electronics, and 2D forms of boron have shown promise in theoretical work. To make a 2D boron crystal monolayer, Guoan Tai at the Nanjing

University of Aeronautics and Astronautics in China and his collaborators deposited boron atoms onto a copper surface.

They showed that the 2D material has electronic and optical properties that could be useful for building devices such as lasers and thin-film photovoltaic panels.

*Angew. Chem. Int. Edn* <http://doi.org/f3jsqh> (2015)



NASA/JPL-CALTECH

## PLANETARY SCIENCE

### How Mars loses its atmosphere

Solar storms have blasted much of Mars's tenuous atmosphere into space over billions of years, making the planet the barren world it is today.

A series of papers has outlined the first results from NASA's Mars Atmosphere and Volatile Evolution (MAVEN) spacecraft, which has been orbiting Mars since September 2014. In one paper, Bruce Jakosky of the University of Colorado Boulder and his colleagues report MAVEN data showing that when protons and electrons from a solar eruption slammed into the planet in March 2015, they increased the rate at which Mars loses its atmosphere by roughly an order of magnitude.

A second paper concludes that only about 10% of the atmospheric particles that leave

Mars are recaptured by its gravitational pull; the rest are lost to space permanently, find David Brain of the University of Colorado Boulder and his co-workers.

And when MAVEN flew within 130 kilometres of the Martian surface, it discovered new populations of charged and neutral particles, including oxygen, nitrogen and carbon dioxide. These were found at unexpectedly low altitudes in the Martian atmosphere, say Stephen Bougher of the University of Michigan in Ann Arbor and his team in a third paper. Together, the findings show that Mars's atmosphere is more complex and dynamic than scientists had thought.

*Science* <http://doi.org/83k> (2015);

*Geophys. Res. Lett.* <http://doi.org/83p> (2015);

*Science* <http://doi.org/83m> (2015)

## REGENERATIVE BIOLOGY

### How mice regrow ear tissue

Blocking a specific cell-signalling pathway in mice boosts the regeneration of ear tissue without any scarring after injury.

Some amphibians and fish can regrow organs and

appendages. To investigate the process in mammals, Thomas Leung, Seung Kim and their colleagues at Stanford University in California studied an engineered mouse model that is adept at regrowing injured ear tissue with no scarring. They found that certain cells in the skin's outer layer produced lower levels of a cell-signalling



molecule called Sdf1 than seen in normal animals. This resulted in reduced recruitment of specific white blood cells that are involved in scar formation.

A compound that inhibits the signalling between Sdf1 and these white blood cells enhanced scar-free regrowth of ear tissue in normal mice. *Genes Dev.* 29, 2097–2107 (2015)

## PHYSICS

## Balloon-popping patterns probed

When balloons burst, they either split open along a single line or, if inflated to a high pressure, shred into many pieces.

Sébastien Moulinet and Mokhtar Adda-Bedia at the École Normale Supérieure in Paris punctured rubber balloons and used a high-speed camera to film the popping process, which lasts for less than 0.1 milliseconds. The stress within the membrane depends on the balloon's internal pressure and on the membrane's thickness and curvature. The authors found that as membrane stress increased, a single crack propagated at higher speeds. Above a threshold level, the crack branched out into as many as dozens of cracks to quickly dissipate the stress, leading to many balloon fragments.

The researchers say that similar principles might cause the tree-like propagation of cracks in processes ranging from the breakup of atomic nuclei to the collisions of asteroids.

*Phys. Rev. Lett.* 115, 184301 (2015)

## MATERIALS

## Ancient art spurs thin batteries

Ink on paper can act as an electrode in a thin, flexible battery.

Inspired by Chinese brush painting, a team led by Xin-Bo Zhang

at the Chinese Academy of Sciences' Changchun Institute of Applied Chemistry fabricated a flexible lithium-air battery using lithium foil and paper with a carbon-based ink (**pictured**). Electrons are stripped from the foil, creating lithium ions that flow to the inked paper electrode, where they combine with oxygen from the air. The resulting battery can hold a charge even after it has been bent 1,000 times. A foldable pack of four batteries, which weighs less than 2 grams, can supply current for 100 hours.

The technique paves the way for cheap and easily manufactured flexible batteries, the authors say. *Adv. Mater.* <http://doi.org/f3js75> (2015)

## CANCER

## Vitamin kills colon-cancer cells

High levels of vitamin C can slow the growth of colorectal tumours in mice.

Lewis Cantley at Weill Cornell Medical College in New York and his colleagues studied human colorectal cancer cells with mutations in either the *KRAS* or the *BRAF* genes, which are commonly mutated in people with this type of cancer. High levels of oxidized vitamin C killed the cells. The authors found that the cancer cells make unusually large amounts of a receptor that

takes up the oxidized vitamin C molecule. Once inside the cell, the compound

kicked off a reaction that led to cell death by depleting a molecule called glutathione, which normally mops up damaging free radicals. High doses of vitamin C stunted colorectal tumour growth in *KRAS*-mutant mice.

With more research,

## SOCIAL SELECTION

Popular topics on social media

### Debate over microbiome initiatives

Researchers studying communities of microbes need to up their game. That was the argument made by two articles published on 28 October in *Science* and *Nature*, which called for national and international initiatives that would unite microbiome researchers and move the field forward. The initiatives would help researchers to develop better, standardized ways of studying microbial communities so that scientists can make meaningful comparisons of data sets across different studies. Some scientists were sceptical, however. Nick Loman, a bacterial geneticist at the University of Birmingham, UK, tweeted: "Just calling for standards and unified data sharing ain't going to make it happen."

But the proponents say that the two articles are just starting points for broader discussion in the field. Microbial ecologist Jack Gilbert at the University of Chicago in Illinois, a co-author of the *Science* article, hopes that the proposals — and the online back and forth — will stimulate further discussion and the creation of new research programmes. "No one is saying that we're going to fundamentally transform the way you do science. We're saying we're going to fundamentally transform the way science is funded and the

way multidisciplinary science can be implemented," he says. "This is starting a conversation."

#### ➔ NATURE.COM

For more on popular papers: [go.nature.com/jugvhd](http://go.nature.com/jugvhd)

*Science* 350, 507–508 (2015); *Nature* 526, 631–634 (2015)



the vitamin could be used to treat some colon cancers, the authors suggest.

*Science* <http://doi.org/83w> (2015)

## ANIMAL BEHAVIOUR

## Fish bring death from above

Needlefish can leap from the sea and dive back down to attack small fish prey from above, the first time such behaviour has been seen in fish.

Ian Tibbetts at the University of Queensland in Australia and his colleagues observed *Tylosurus gavioloides* needlefish (**pictured**) pursuing prey that were in tightly packed shoals in two locations off the Queensland coast. Needlefish

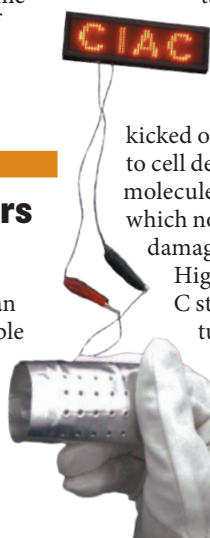
normally strike at prey found within about 50 centimetres of their body, but aerial strikes extended their range to more than 2 metres. And such air strikes prevented fish prey from fleeing to the surface and leaping to escape, which they do when they are attacked from below the surface of the water.

The authors suggest that needlefish could be exploiting an optical effect called Snell's window — which limits vision through the water's surface — to mask their approach.

*J. Fish Biol.* <http://dx.doi.org/10.1111/jfb.12799> (2015)

#### ➔ NATURE.COM

For the latest research published by Nature visit: [www.nature.com/latestresearch](http://www.nature.com/latestresearch)





# SEVEN DAYS

The news in brief

## EVENTS

### Science unmuzzled

Canadian scientists have been thrilled by their new Liberal government's first week in office. On 5 November, the government reinstated the mandatory long-form population census, which scientists and policymakers rely on for social data. And on 6 November, the new Minister of Innovation, Science and Economic Development, Navdeep Bains, 'unmuzzled' government scientists, declaring that they no longer need to seek permission from communications officers to speak with the media or public. Prime Minister Justin Trudeau was sworn in on 4 November and announced Kirsty Duncan in the new post of science minister (see page 146 for more).

### Ebola-free nations

The World Health Organization (WHO) declared the end of the Ebola outbreak in Sierra Leone on 7 November. On that date, the country met the WHO's criteria of 42 days — twice the maximum incubation period for the virus — since the last case was cured. Liberia was declared Ebola-free on 3 September, leaving Guinea as the only remaining country in the region with continuing infections. One new case of Ebola was recorded in Guinea in the week up to 1 November, and three infections were reported in the previous week. The WHO warns that there is a risk of further cases to come.

### Amazonian park

Peruvian President Ollanta Humala announced the creation of the Sierra del Divisor National Park on 8 November, a move long sought by scientists and conservationists. Encompassing 1.3 million hectares, the park protects



LOS ALAMOS NATL LAB./SPL

## Manhattan Project gets national park

The US Department of Energy and the National Park Service established a national park on 10 November to preserve historic sites from the 1940s Manhattan Project to develop atomic bombs. The park covers three sites — Oak Ridge in Tennessee; Hanford in Washington; and Los Alamos in New Mexico. Each site worked on different aspects of nuclear-weapons design and

uranium and plutonium production. Preserved facilities include Oak Ridge's X-10 graphite reactor, which converted uranium-238 to plutonium-239; Hanford's B reactor, the first to produce large amounts of plutonium; and the V-Site buildings in Los Alamos, where components for the first nuclear bombs were assembled (pictured: the Trinity bomb).

crucial Amazonian biodiversity, including endangered plant and animal species, and is also home to uncontacted indigenous groups. The philanthropic Andes Amazon Fund in Washington DC called the decision "one of the most significant rainforest conservation accomplishments in recent years". Sierra del Divisor faces mining, logging and other threats that officials must continue to fend off.

## BUSINESS

### Drug-cost probe

The US Senate is investigating drug costs and price hikes resulting from mergers

and acquisitions. On 4 November, two senators requested documents from four pharmaceutical companies relating to their communication with regulators and details on drugs that they acquired from other companies. Central to the investigation is Turing Pharmaceuticals in New York City, which caused outcry in September when it acquired the antiparasitic drug pyrimethamine (Daraprim) and increased its price by more than 5,000%. Turing's chief executive, Martin Shkreli, said this week that the firm will most likely drop the price by about 10% before the end of the year.

### Oil giant quizzed

New York state attorney-general Eric Schneiderman has launched an investigation into whether ExxonMobil deliberately misled investors about the risks of global warming. Schneiderman issued a subpoena on 4 November for records related to the oil company's statements about climate change. Environmentalists have accused ExxonMobil, headquartered in Irving, Texas, of intentionally lying about the science underlying global warming and of funding climate-change sceptics to prevent the regulation of greenhouse gases. Company officials deny

the claims, saying that the firm supports the implementation of a carbon tax.

## VW emission woes

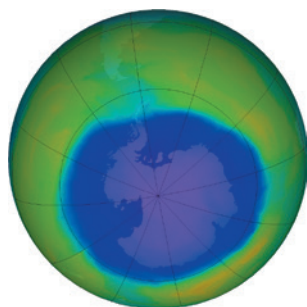
After an internal investigation, Volkswagen (VW) reported on 3 November that 800,000 of its cars show “irregularities” in fuel-consumption rates and carbon dioxide emissions, which may be different from those officially stated. The report comes on top of the company’s manipulation of nitrogen oxide tests in a scandal involving 11 million VW cars worldwide.

Environmental groups claim that real-world CO<sub>2</sub> emissions from cars in the European Union are up to 40% higher than the official values that manufacturers report based on idealized laboratory tests.

### POLICY

## HFC curbs

Governments agreed on 6 November to use the Montreal Protocol ozone treaty to regulate the climate impact of hydrofluorocarbons (HFCs), commonly used as refrigerants. HFCs were developed to replace chemicals that deplete the ozone layer (**pictured** is the ozone hole over the Antarctic in October this year), but they are powerful greenhouse gases. At a meeting in Dubai, 197 parties to the treaty agreed to manage HFCs under the



protocol, but negotiators delayed debate about the details of any regulations until 2016.

### PEOPLE

## Obokata PhD blow

Waseda University in Tokyo has revoked the doctorate of Haruko Obokata, formerly of the RIKEN Center for Developmental Biology in Kobe, Japan. Obokata was lead author on two papers published in *Nature* in January 2014 that reported the creation of pluripotent stem cells by subjecting body cells to acid; the method was called stimulus-triggered acquisition of pluripotency (STAP). Problems in the papers led to their retraction in July 2014. In October that year, after finding plagiarized passages and images in her 2011 PhD dissertation, the university gave Obokata a year to make corrections. But in its 2 November statement, the university committee judged her drafts inadequate.

## Euro advisers

The seven researchers who will form a panel to provide the European Commission with independent scientific advice were named on 10 November. They include Rolf Dieter-Heuer, the outgoing head of CERN, Europe’s particle-physics laboratory near Geneva, Switzerland, and Julia Slingo, chief scientist at Britain’s Met Office in Exeter. The panel’s announcement marks the formal launch of the commission’s revamped system for obtaining scientific advice. Under that system, the panel will feed in advice to the commission by seeking input from national academies and the wider scientific community. See [go.nature.com/7gj4pu](http://go.nature.com/7gj4pu) for more.

## Prize scoop

On 10 November, *Nature* reporter Alexandra Witze took home the gold prize for magazine journalism at the 2015 AAAS Kavli Science Journalism Awards. The awards, which cover magazine, newspaper and television journalism, are funded by the Kavli Foundation and administered by the American Association for the Advancement of Science. The judges singled out Witze’s features for praise, including ‘The quake hunters’ (*Nature* 523, 142–144; 2015) and

## COMING UP

### 16–20 NOVEMBER

Physicists meet in Paris to celebrate 100 years of general-relativity theory. [go.nature.com/phmn3r](http://go.nature.com/phmn3r)

### 18–22 NOVEMBER

AORTIC 2015, the 10th International Conference on Cancer in Africa, is held in Marrakesh, Morocco. [aorticconference.org/](http://aorticconference.org/)

‘The Pluto siblings’ (*Nature* 518, 470–472; 2015).

### FUNDING

## Indian alliance

India’s ministry of human-resource development launched a research alliance between 16 Indian Institutes of Technology (IITs) and the Indian Institute of Science (IISc), Bangalore, on 5 November. The Impacting Research Innovation and Technology (IMPRINT) programme is the country’s first pan-IIT and IISc initiative. Its ten topics range from water resources and river systems to nanotechnology. The initiative has an initial budget of 10 billion rupees (US\$150 million).

## Google heart boost

On 8 November, Google’s life-sciences branch and the American Heart Association announced that each would contribute US\$25 million to a \$50-million project to make breakthroughs in understanding, treating and preventing coronary heart disease, which causes one in every three US deaths. The money will all go to one interdisciplinary team, led by a single investigator and advised by an as-yet unchosen panel. Applications will open in January.

➔ [NATURE.COM](http://NATURE.COM)

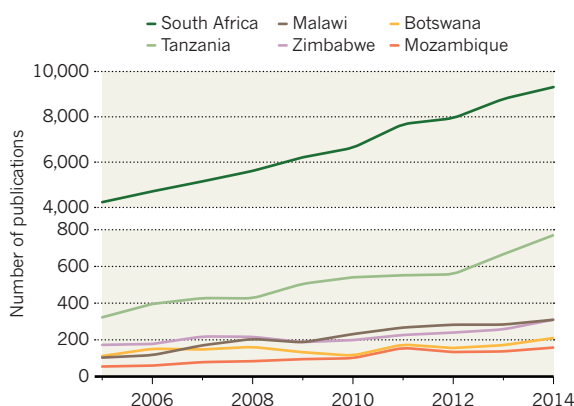
For daily news updates see: [www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

The latest *UNESCO Science Report* highlights the striking progress made by sub-Saharan Africa in science, technology and innovation. Publications by nearly all 15 countries in the Southern Africa Development Community rose from 2005 to 2014. Article output almost tripled in Malawi and Mozambique, and more than doubled in South Africa and Tanzania. The 10 November report points to Malawi’s huge effort to attract foreign investment and increase science spending to offset its lack of natural resources.

## THE RISE OF SOUTHERN AFRICAN SCIENCE

Since 2005, the number of publications from southern African nations has risen dramatically, almost tripling in Mozambique and Malawi.





# NEWS IN FOCUS

**UK POLITICS** Research jitters ahead of Conservative spending review **p.144**

**TELESCOPES** Giant Arecibo observatory faces murky future **p.142**

**MEGA PRIZE** More than 1,000 physicists split US\$3-million award **p.145**



**SOCIAL MEDIA** How hashtags can help to fight sexism in science **p.148**

TOM PENNINGTON/GETTY



US President Barack Obama's decision to cancel the Keystone XL oil pipeline was nearly seven years in the making.

## ENERGY

# Pipeline rejection puts US climate politics in spotlight

*President Barack Obama abandons Keystone XL oil pipeline from Canada.*

BY JEFF TOLLEFSON

US President Barack Obama's rejection on 6 November of the proposed Keystone XL oil pipeline may do little to reduce greenhouse-gas emissions directly. But the long-awaited decision — a symbolic victory for environmental groups — signals Obama's growing ambition to combat climate change despite strong political opposition.

"America is now a global leader when it comes to taking serious action to fight climate change," he said at the White House. "And frankly, approving this project would have undercut that global leadership."

The 1,900-kilometre pipeline was to have carried crude oil from the tar sands of Alberta, Canada, to existing pipelines in the US Midwest that run to refineries along the Gulf of Mexico. The project, backed by pipeline firm

TransCanada, was controversial in part because cumulative greenhouse-gas emissions from tar-sands oil are up to 20% higher than those from conventional crude oil, on average.

Obama's announcement comes less than a month before the start of United Nations Paris climate talks, which are expected to produce a new international agreement to combat global warming. The timing pleases environmentalists, who have pushed the pipeline ▶

► to the top of the US political agenda with a sustained campaign that included public demonstrations at the White House.

“Rejecting the pipeline makes it tougher to dig up tar sands that would only add more fossil fuels to the fire,” said Lou Leonard, vice-president for climate change at the conservation group WWF in Washington DC, in a statement. “We hope it continues the momentum we’re seeing to ratchet up climate ambition for Paris and beyond.”

Republican lawmakers have argued just as vociferously in favour of the pipeline, accusing environmentalists of trying to kill jobs and drive up energy prices. Keystone XL vaulted into the US presidential campaign agenda in 2012, and its spectre will haunt next year’s race to pick Obama’s successor. Several Republican presidential candidates quickly vowed to reverse Obama’s decision on the pipeline if given the chance.

“When I’m president, Keystone will be approved, and President Obama’s backwards energy policies will come to an end,” tweeted Republican contender Marco Rubio, a Senator from Florida.

Obama addressed his critics head on in his White House speech, arguing that there is no economic justification for approving Keystone XL. Oil production in the United States is at its highest level in years, he said, and oil prices have fallen sharply. Obama also

noted that the US economy is growing even as greenhouse-gas emissions decline — due in part to his administration’s regulations for curbing vehicle emissions of carbon dioxide.

Some industry analysts argue that the cancellation of Keystone XL will have a minimal impact on the tar sands’ overall greenhouse-gas output, however, because energy companies can merely ship their product to market by rail or through other pipeline projects that are already in the works.

“Keystone was a very prominent project given the symbolic nature it took on, but there are a lot of projects and options out there,” says James Burkhard, vice-president of the consultancy IHS Energy in Washington DC.

His firm estimates that oil production from the tar sands could increase by up to 25% by 2020 compared to the 2014 level, based on projects that are already approved and under construction. The price of oil exerts a strong influence over oil-sands investment, Burkhard says — but the cancellation of Keystone XL boosts uncertainty about tar-sands development after 2020.

And there is some evidence that this uncertainty is already having an effect: on 27 October,

energy giant Shell cancelled its Carmon Creek tar-sands extraction project, citing the difficulty of shipping Canadian oil to markets. The energy firm Statoil voiced similar concerns when it cancelled another tar-sands venture last year.

Shell is not abandoning Alberta altogether, however. Hours after Obama rejected the pipeline, the company formally launched a project to capture and store more than 1 million tonnes of CO<sub>2</sub> each year at an oil-processing plant in the province. But the Can\$1.35-billion (US\$1-billion) project relies heavily on subsidies, and there is as yet no plan to make such projects economically viable in the future, says Simon Dyer, regional director for Alberta at the Pembina Institute, an environmental think-tank in Calgary.

Dyer says that Obama’s decision on Keystone XL could help to reshape Canada’s climate policy. The country’s new prime minister, Justin Trudeau, is a supporter of Keystone XL who has also pledged to combat climate change. And Alberta’s new premier, Rachel Notley, plans to release parts of a climate policy for the province before the UN Paris meeting begins.

If Canada wants to develop markets for its oil, Dyer says, it has to create a credible climate policy and decide how much of its carbon budget should go to the tar sands. “This is an extremely controversial and polarizing thing to say in Canada,” Dyer says. “The conversation is long overdue.” ■

**“Keystone was a very prominent project ... but there are a lot of projects and options out there.”**

## ASTRONOMY

# Arecibo Observatory head quits after funding row

*Departure of long-term advocate adds to the woes of the financially troubled radio telescope.*

BY TRACI WATSON

Physicist Robert Kerr uses irony to describe the first hint of trouble: “Radio quiet,” he calls it. After four years as director of the Arecibo Observatory, home to the world’s largest single-dish radio telescope, he says that he was suddenly out of the loop: contacts at both the US National Science Foundation (NSF), which owns the observatory, and SRI International, the contractor that runs it, stopped returning his e-mails and phone calls.

After a month of silence, Kerr was stripped of his role as the observatory’s principal investigator. Shortly afterward, he resigned from his other post, as operations director.

Kerr traces his departure to a disagreement

over a possible windfall for the Puerto Rico observatory. In late July, he publicly criticized the NSF for planning to cut its contribution to Arecibo if the facility began to take payments for helping in a private survey looking for signs of extraterrestrial intelligence. NSF officials say that his assertions are inaccurate and that his communication with Kerr never lapsed.

Whatever the facts, some Arecibo observers see Kerr’s exit as an ill-timed loss for a storied, but financially threatened, scientific facility that faces a murky future.

“Somebody’s going to have to be the person actively trying to figure this out,” says Michael Nolan, a former Arecibo director now at the University of Arizona in Tucson. “Bob was that person. Without him, I don’t know what their plan is.”

The drama surrounding Kerr’s departure is in keeping with the scale of Arecibo, which has a bowl-shaped reflector that measures 305 metres across and is the world’s most sensitive radio telescope. At Arecibo, researchers made the first discovery of a binary pulsar — the 1974 find won a Nobel prize in 1993 — and of the first planets outside our Solar System.

Today, 52 years after it began operations, Arecibo is still one of the world’s go-to telescopes for getting a close look at potentially hazardous asteroids, and the facility remains a key tool for studying pulsars and Earth’s upper atmosphere.

But past glories and present capabilities may not ensure its survival. The NSF, which provides two-thirds of the observatory’s US\$12-million annual budget, is strapped for cash





The Arecibo Observatory in Puerto Rico is beset by budget problems.

to build and operate new telescopes that are high priorities for the astronomy community, such as the Large Synoptic Survey Telescope now under construction in Chile. In 2006, an expert panel recommended that the agency close Arecibo unless someone else could be found to foot the bill. NASA began kicking in with money five years ago and now contributes \$3.7 million annually, but so far no one has materialized to pay the rest.

Meanwhile, the NSF faces pressure to keep grant money flowing while it funds new telescopes. A 2012 expert report warned that unless the NSF slashes the amount it spends on large facilities such as Arecibo, research grants to astronomers could be “decimated”.

An avowed Arecibo champion, Kerr was the observatory’s on-site director from 2007 to 2008 and returned in 2011, taking up the roles of both operations director and principal investigator. During his tenure, proposals for the use of Arecibo rose, he says, and a system for studying the ionosphere with high-frequency

radio waves was rebuilt. Even so, budget cuts forced the lay-off of some 20% of the staff a decade ago. The facility is afflicted with “stagnation”, says radio astronomer Alex Wolszczan at Pennsylvania State University in University Park, who in 1991 used Arecibo to make the first detection of extrasolar planets.

Kerr, an upper-atmosphere physicist, made the best of a difficult and frustrating job, colleagues and users say. “Bob really cared about the observatory, and he really wanted to find a way to make it work,” says University of Arizona planetary scientist Ellen Howell. She and her husband Michael Nolan were researchers at Arecibo until this summer.

About that time, a potential Arecibo saviour appeared in the form of Russian billionaire Yuri Milner. Milner’s Breakthrough Listen project is funding a \$100-million effort to search for intelligent extraterrestrial life, and wanted to enlist Arecibo’s help.

In Kerr’s telling, NSF officials told him that if Arecibo got funding from Breakthrough, its

own funding would fall by the same amount. In a 29 July article, an angry Kerr told the magazine *Scientific American* that the NSF had placed Arecibo in an “unscrupulous” bind: walk away from the Breakthrough money, or accept it and lose NSF dollars.

But NSF officials say that it was not like that. “It was expected that some offset would occur because this situation would divert telescope time” away from other science, says NSF astronomy-division director Jim Ulvestad. But the agency has still not decided whether Breakthrough funding would trigger a one-for-one cut or indeed any cut at all, he says. Kerr and SRI were both told so repeatedly, Ulvestad says.

## COMMUNICATION BREAKDOWN

It was after the *Scientific American* article appeared, Kerr says, that communication with his superiors at SRI and with NSF officials ceased almost entirely. He got an e-mail notice a month later that he was no longer Arecibo’s principal investigator and, feeling hamstrung, decided to step down from the operations director’s job as well. He does not regret speaking out, but “I certainly regret that I’m no longer a primary advocate for the observatory”.

The NSF says that communication with Kerr continued as usual, and Kerr concedes that his regular biweekly phone calls with the agency did not end. The NSF referred other questions about Kerr’s tenure to his former employer, SRI, which says that it does not discuss matters of personnel.

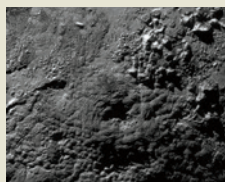
Just after Kerr cleared his desk, the NSF appealed in a ‘Dear colleague’ letter for proposals to run the observatory, calling especially for ideas that “involve a substantially reduced funding commitment from NSF”. The agency says that it is now reviewing whether it will continue to support the facility. Kerr, who worked to find private parties to manage and fund the observatory, says he remains hopeful that a coalition of universities and foundations will emerge to rescue Arecibo.

But others are less optimistic. Efforts to find an outside donor have been ongoing for some time, to no avail, Nolan says. Users can be expected to pay for telescope time, but “someone has to pay for the base operations — keeping the grass cut, keeping the big steel structure from falling out of the sky”, he says. “And that’s the part everyone’s finding too expensive.” ■ **SEE EDITORIAL P.134**

JAY M. PASACHOFF/SCIENCE FACTION/CORBIS



## TOP STORY



Icy volcanoes may dot Pluto’s surface  
[go.nature.com/xqvouy](http://go.nature.com/xqvouy)

## MORE NEWS

- Psychologists’ betting market hints at most reliable research findings  
[go.nature.com/sjhp19](http://go.nature.com/sjhp19)
- Light and heat propel bendy graphene crawlers  
[go.nature.com/nw3uez](http://go.nature.com/nw3uez)
- A glaciologist’s account of life in Antarctica  
[go.nature.com/twfl5s](http://go.nature.com/twfl5s)

## NATURE PODCAST



Sexism in science on Twitter; porous liquids; and the relationship between humans and bees  
[nature.com/nature/podcast](http://nature.com/nature/podcast)

NASA/JHUAPL/SWRI



Britain's chancellor George Osborne (centre) will announce a spending review on 25 November.

## POLITICS

# Spectre of change haunts UK science

*Scientists fear flat budget ahead of the Conservative government's first spending review.*

BY DANIEL CRESSEY & ELIZABETH GIBNEY

Uncertainty is haunting UK science. On 25 November, the Conservative government that won a parliamentary majority in May will announce its first spending review — and researchers have no idea what to expect.

Many hold that the science budget desperately needs to increase, and there is renewed pressure to implement a rise from an influential group of politicians. Current investment from both business and government is well below 2% of gross domestic product (GDP). That should be increased to 3%, an internationally recognized target, to safeguard productivity, warned the House of Commons committee on science and technology on 9 November, following an investigation into research spending.

"We need to have that increase," says Nicola Blackwood, the Conservative Member of Parliament who heads the committee. "We need a road map to the 3% target."

But despite pleas for more cash, the goal of the spending review — to reduce

overall government outgoings by £20 billion (US\$30 billion) by 2020 — does not give most researchers hope. If the science budget is part of the trimming — or if it remains flat, as occurred in the previous comprehensive spending review in 2010 — many say that science will suffer.

"We are so lean now I really can't take anything else out," says Michael Wakelam, director of the Babraham Institute, a life-sciences research centre near Cambridge. Even a flat budget would mean that he could not replace staff who leave and that the quality of research at the institute would decline, he says.

Intensifying the uncertainty is a government proposal announced on 6 November to axe the Higher Education Funding Council for England, which doles out around £1.6 billion each year to universities for research; scientists are also awaiting the results of a government-commissioned review into the seven UK research councils, which dish out a further £3 billion.

The 2010 spending review kept the core science budget frozen at £4.6 billion a year for five years, and slashed a smaller budget for science

infrastructure by almost half.

The result was a relief to researchers because the government — then a coalition between the Conservatives and the Liberal Democrats — had committed to substantial cuts in public spending ahead of the review. David Willetts, science minister at the time, even received flowers from the founder of *Research Fortnight* magazine in the wake of the announcement, for his role in protecting science.

Although cash for infrastructure has increased since then, inflation has eroded the core science budget, forcing researchers to cut back. Another five years of the same would now be extremely difficult, says Alison Davenport, who is chair of the independent scientific advisory board to the Science and Technology Facilities Council, and provided evidence about the impact of frozen science funding to Blackwood's parliamentary committee.

Another freeze on science funding "will lead to an accelerated decline with serious and irreversible damage to particle physics", Davenport warned the committee. In one scenario that she considered, Britain's flagship X-ray synchrotron, the Diamond Light Source, would operate at 70% of its optimum capacity, the national Central Laser Facility at 50%, and the ISIS neutron source would need to shut down by 2019.

There are also "serious concerns", says Davenport, that there will not be enough resources to analyse the data coming in from major astronomy projects and the Large Hadron Collider.

The government, however, may not appreciate the impact of a flat budget this time, says Nick Hillman, director of the Higher Education Policy Institute in Oxford, particularly because people regarded it as positive in 2010.

Even worse than a frozen budget would be cuts. That fear has been stoked by the government's demand in July that before the spending review, almost all departments — including that for Business, Innovation and Skills, which manages most of the government-supported science budget — model scenarios in which their budgets are cut by 25% and 40%.

Few think that cuts to the science budget would reach even 25%. Still, at least one major funder, the Medical Research Council, believes that there is a risk that its funds could be cut and has examined various possible scenarios, according to the minutes of a July meeting.

Worried researchers have started taking action: more than 1,000 have written to the government through the Science is Vital campaign. Astrophysicist Phil Evans of the University of Leicester warned in his message that skilled academics are leaving the country because they can not find jobs; others detailed how their own research contributes to the UK economy.

'Backdoor' cuts, says Kieron Flanagan, a science-policy researcher at the Alliance Manchester Business School, could include stretching the science budget to cover areas such as innovation, which is currently separate, or cutting the research



councils' administrative budgets.

Further uncertainty comes in the form of the characters behind the review. The highly respected Willetts is gone, and the current science minister, Jo Johnson, was notably unimpressed by the notion that Britain should be spending 3% of its GDP on research, describing it as "a nice round number, more than anything else" when he appeared before Blackwood's committee.

Chancellor of the Exchequer George

Osborne, who holds the purse strings and is seen as the architect of the government's economic austerity programme, has consistently claimed to support science. Osborne got kudos for sparing science in the last review — and for increasing research-infrastructure spending since then. "It would be very odd to spend your first five years as chancellor saying, 'I'm the chancellor for science,'" says Hillman, "and then to not see that through in the second term."

Any science spending boost might come

with strings attached. Innovation and regional growth are likely to be priorities of future science spending, says Paul Nightingale, deputy director of the Science Policy Research Unit at the University of Sussex in Brighton. So Osborne might continue a trend, established under the last government, of allocating funding directly from the treasury to projects outside London, such as the UK National Graphene Institute at the University of Manchester, in an effort to boost regional economies. ■

## PHYSICS

# Mega science prize split between more than 1,000 physicists

*Multimillion-dollar Breakthrough awards announce winners in glitzy ceremony.*

BY ZEEYA MERALI

**I**t is a celebration of the collective over the individual. Whereas only two physicists picked up this year's physics Nobel prize for the discovery that neutrinos have mass and can change identity while travelling, 1,377 collaborators who were involved in the 'neutrino oscillation' experiments behind the finding will share the US\$3-million Breakthrough Prize in Fundamental Physics.

The awards, announced on 8 November at a star-studded ceremony at NASA's Ames Research Center in Moffett Field, California, also honour five biologists in the life sciences and a mathematician.

"There is a message here that science is a much more collective effort than it was 100 years ago," says Russian Internet entrepreneur Yuri Milner, who is one of the prizes' founders. "It is international, it is diverse, it involves lots of people."

The Breakthrough physics prize marks the first time that a significant science prize has been awarded to such a large group of people — which Milner describes as "a logistical nightmare". It recognizes the members of five international experiments that established that neutrinos have mass, which contradicts the standard model of particle physics. In October, the Nobel prize was shared between just two people for the same discovery: Arthur McDonald at Queen's University in Kingston, Canada, and Takaaki Kajita at the University of Tokyo. (Both are among the Breakthrough prizewinners.)

"This is recognition for excellent science that could only be achieved by cooperation between many scientists," says McDonald, who led an experiment at the

Sudbury Neutrino Observatory in Canada.

Each of the five teams will receive \$600,000; team leaders are allotted two-thirds of the money, with the remaining one-third split between other team members.

Because it follows hot on the heels of the Nobels, the Breakthrough prize could be seen as a direct criticism of its older rival, which notoriously refuses to honour more than three individuals in science categories. But Edward Witten, a physicist at the Institute for Advanced Study in Princeton, New Jersey, who chaired the Breakthrough selection committee, notes that the winners were chosen over the summer, before the Nobels were revealed, so the overlap is coincidental. Nonetheless, Witten says, the decision makes a deliberate statement: "We ... consider it important to include at least at a symbolic level the many scientists who contributed."

Göran Hansson, former secretary of the Nobel Committee for Physiology or Medicine, stands by the Nobel-prize policy, however. "Precisely because there is so much emphasis on huge organizations, we feel it is important to identify the individuals who pioneered the discoveries," he says. The Breakthrough prizes are no threat to the prestige of the Nobel, Hansson adds: "We have more than a century of legacy, and people will continue to look to the Nobels to identify excellence in science."

Five biologists were also awarded Breakthrough prizes. Neuroscientists Karl Deisseroth at Stanford University in California

and Ed Boyden at the MIT Media Lab in Cambridge, Massachusetts, received separate awards for developing optogenetics — the programming of neurons so that their electrical activity can be controlled by light. Helen Hobbs at the University of Southern Texas Medical Center in Dallas was recognized for her discovery of human genetic variants that alter cholesterol levels. John Hardy at University College London was honoured for finding the mutations in the gene encoding amyloid precursor protein that cause early-onset Alzheimer's disease, and Svante Pääbo at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, was recognized for sequencing ancient genomes.

Mathematician Ian Agol at the University of California, Berkeley, received an award for proving three conjectures relating to how 3D 'manifolds' (higher-dimensional equivalents of two-dimensional surfaces) can be flattened and transformed. The work could one day have applications for understanding how space-time curves, says Agol, who has previously received the Clay prize and other established maths prizes.

Compared to those awards, says Agol, the Breakthrough prize is too new to have established a clear status, and does not yet have the same level of fame among mathematicians. "In any case, it is a great honour to be receiving this," he adds — although he admits that he didn't relish appearing on television.

The award ceremony, hosted by *Family Guy* creator Seth MacFarlane, was broadcast live for the first time on the National Geographic Channel, and an edited version will be aired on FOX. It also honoured eight early-career scientists, who received \$100,000 each, and the winner of a prize for schoolchildren. ■

## GOVERNMENT

## Canadian leader names science ministers

*New prime minister gives research a higher profile.*

BY NICOLA JONES

Canada's new prime minister, Justin Trudeau, took office on 4 November — and as one of his first acts, created the post of Minister of Science.

Kirsty Duncan, a medical geographer at the University of Toronto in Canada, will be the first to hold the job. Duncan, who contributed to the 2001 report of the Intergovernmental Panel on Climate Change, has also written a book about the 1918 Spanish flu epidemic.

Her appointment marks a change from the government of former prime minister Stephen Harper. His administration placed oversight of science in the hands of a junior minister of state in the Industry Canada department.

"Harper collapsed the purview of science into the purview of industry," says Carol Linnitt, an environmental-policy analyst at the Vancouver-based non-profit environmental group DeSmog Canada.

Scientists and science groups say that they are excited by Duncan's appointment but want to know more about the Minister of Science's responsibilities. "A real minister! And someone with a PhD!" says Marc Saner, former director of the Institute for Science, Society and Policy at the University of Ottawa. "From the point of view of image, it's great. How this works in practice, I don't know."

Trudeau has also appointed Navdeep Bains, a financial analyst, as Minister of Innovation, Science and Economic Development.

"If we take it at face value, we now have two ministers responsible for science," says Rees Kassen, a biologist at the University of Ottawa and chair of the Partnership Group for Science and Engineering, an Ottawa-based association of science and engineering organizations. He suspects that Duncan will work to ensure that the government conducts research in areas that universities and businesses are not exploring, whereas Bains will seek to encourage technological innovation in the private sector.

Trudeau, who is expected to appoint a chief science adviser, has also picked an environment minister — renaming the post as Minister of Environment and Climate Change. The new minister, Catherine McKenna, is a lawyer whose work has focused on international trade, investment and constitutional issues. ■



SHARON LEES/GREAT ORMOND STREET HOSPITAL

Layla Richards is in remission from leukaemia after getting cells treated with DNA-cutting enzymes.

## GENETIC MODIFICATION

## Gene-editing wave hits clinic

*Companies prepare to test range of therapies in people.*

BY SARA REARDON

Layla Richards, a one-year-old girl with leukaemia, is in remission thanks to gene-editing technology that allowed her to receive modified immune cells from another person.

Her case is the second use of gene-editing as a treatment in humans — the first was last year in people with HIV. Similar trials are planned, and companies are also preparing to test therapies that inject DNA encoding gene-editing enzymes directly into the human body.

The team that treated Richards, led by immunologist Waseem Qasim of the Great Ormond Street Hospital for Children NHS Trust in London, had planned to start a safety trial of the gene-editing technology in 10–12 people next year. But when the researchers came across the baby, in whom all other treatments had failed, they obtained special permission to treat her. Several months on, Qasim says that she is doing well. They will present her case in December at an American Society of Hematology meeting in Orlando, Florida.

To administer the therapy, the researchers extract immune cells called T cells from a healthy donor and expose them to DNA-cutting enzymes known as TALENs. The system, developed by researchers from Cellectis,

a company headquartered in Paris, is designed to deactivate immune genes that would otherwise cause the donor cells to attack when injected into another person. The system also modifies genes to protect the cells from anti-cancer drugs. The individual then undergoes therapy to destroy his or her own immune system, which is replaced with the modified cells. The treatment is not a permanent solution for leukaemia, says Qasim, but rather a 'bridge' to keep someone alive until a matched T-cell donor can be found.

## HIV SUCCESS

The first ever use of gene-editing in people employed a similar *ex vivo* approach. Last year, Sangamo BioSciences in Richmond, California, published results from a clinical trial in which it used gene-edited cells to treat 12 people with HIV (P. Tebas *et al.* *N. Engl. J. Med.* 370, 901–910; 2014). Instead of TALENs, the researchers used a DNA-cutting enzyme called a zinc-finger nuclease (ZFN).

When added to blood extracted from patients, the ZFNs cut out the gene for a protein on T cells targeted by HIV, and the team then pumped the cells back into the patients. The results were positive — at the time of the announcement, half of the participants had been cleared to stop taking their antiretroviral



drugs. Sangamo tells *Nature* that it has now treated more than 70 people with the therapy.

For some diseases, however, it makes more sense to edit the genome *in vivo* — for example, if the target cells are in an organ or tissue type that is harder to remove than blood.

In a study presented in October at a meeting of the US National Academies of Sciences, Engineering and Medicine in Washington DC, Sangamo senior scientist Fyodor Urnov reported that his group had injected 15 monkeys with viruses that carried genes encoding a ZFN and normal versions of factor IX — a blood-clotting protein produced in the liver that is mutated in people with haemophilia B.

The ZFN cut the genome at a section that encodes a protein called albumin, which is made in large quantities in the liver, and inserted a healthy version of the factor-IX gene. The monkeys began producing much more factor IX: levels in the blood increased by 10%. Urnov says that the albumin site could be a good place to insert other genes, likening it to “a USB port in the human genome” (R. Sharma *et al. Blood* **126**, 1777–1784; 2015).

A committee at the US National Institutes of Health, which approves all clinical trials involving modified DNA, gave the green light to human trials of the factor-IX therapy in September, according to Urnov, but Sangamo must still get permission from the US Food

and Drug Administration (FDA). Urnov says that the company will apply by the end of the year, and that trials could begin in early 2016. Sangamo plans to apply for permission to do several other trials of gene-editing *in vivo*, including of therapies for the blood diseases haemoglobinopathy and  $\beta$ -thalassaemia.

Others also plan to start testing the approach in people. On 3 November, biotechnology start-up Editas Medicine in Cambridge, Massachusetts, announced that, by 2017, it hopes to start trials of *in vivo* gene-editing. The researchers would inject DNA encoding the CRISPR/Cas9 enzyme system into the eyes of people with a rare retinal disorder called Leber congenital amaurosis to correct a mutated gene.

Therapies both *ex vivo* and *in vivo* risk causing cuts and mutations elsewhere in the genome, but the *in vivo* scenario introduces another concern because the DNA-delivering vector can remain active in the body for years after injection. This could have unforeseen effects, such as inducing an immune reaction to the DNA-cutting enzyme, worries biologist Valder Arruda of the University of Pennsylvania in Philadelphia, who is exploring treatments

for haemophilia that involve conventional gene therapy. Sangamo, however, says that it has not seen evidence of such effects in its animal studies. Other challenges of editing *in vivo*, says Qasim, include ensuring that enough of the target cells actually get edited and that the vector delivers its payload to the right part of the body.

The list of disorders that editing *in vivo* might help to treat is growing. At a synthetic biology meeting in April, biomedical engineer Charles Gersbach at Duke University in Durham, North Carolina, reported a study done in mice with the mutation responsible for muscular dystrophy, a muscle-wasting disease. When the team injected a viral vector encoding a DNA-cutting enzyme into the mice's muscles, the injections corrected the gene in about 20% of muscle cells — enough to improve tone and strength substantially. “I think [*in vivo*] is the next wave of gene editing,” says Gersbach. ■

**The treatment is a ‘bridge’ to keep someone alive until a matched T-cell donor can be found.**

#### CLARIFICATION

In the World View ‘Forensic DNA evidence is not infallible’ (*Nature* **526**, 611; 2015), it may not have been clear that the current investigation in Texas is reportedly focusing on statistics and not the specific problem of secondary contamination in touch-DNA samples.

# WE CAN # IT

BY LAUREN MORELLO





# Social media is shaking up how scientists talk about sexism and gender issues.

**W**hen Fiona Ingleby took to Twitter last April to vent about a journal's peer-review process, she didn't expect much of a response. With only around 100 followers on the social-media network, Ingleby — an evolutionary geneticist at the University of Sussex near Brighton, UK — guessed that she might receive a few messages of support or commiseration from close colleagues. What she got was an overwhelming wave of reaction.

In four pointed tweets, Ingleby detailed her frustration with a *PLoS ONE* reviewer who tried to explain away her findings on gender disparities in the transition from PhD to postdoc. He suggested that men had “marginally better health and stamina”, and that adding “one or two male biologists” as co-authors would improve the analysis. The response was a full-fledged ‘Twitterstorm’ that spawned more than 5,000 retweets, a popular hashtag — #addmaleauthorgate — and a public apology from the journal. “Things went really mental,” Ingleby says. “I had to turn off the Twitter notifications on my e-mail.” Yet her experience is not as unusual as it may seem.

Social media has enabled an increasingly public discussion about the persistent problem of sexism in science. When a male scientist with the European Space Agency (ESA) wore a shirt patterned with half-naked women to a major media event in November 2014, Twitter blazed with criticism. The site was where the first reports surfaced in June of Nobel Prizewinning biologist Tim Hunt's self-confessed “trouble with girls” in laboratories. And in mid-October, many astronomers took to Twitter to register their anger and disappointment when the news broke that Geoffrey Marcy, an exoplanet hunter at the University of California, Berkeley, was found to have sexually harassed female subordinates for at least a decade.

“I have been in [the] field for 15 years,” wrote Sarah Hörst, a planetary scientist at Johns Hopkins University in Baltimore, Maryland. “It is my field now too & we are not going to do things this way anymore if I have anything to do w/ it.”

Scientists studying the rise of social media are still trying to understand the factors that can whip an online debate into a raging Twitterstorm. Such events often have far-reaching and unpredictable consequences — for participants as well as targets. Sometimes this continuing public discussion prompts action: *PLoS ONE* is re-reviewing Ingleby's paper, and its original editor and reviewer no longer work for the journal, for example. But women who speak out about sexism often face a vicious backlash, ranging from insults to threats of physical violence.

Although it is not yet clear whether the social-media conversation about sexism in science will help to create lasting change, some scientists think that it may provide a sense of solidarity for women across disciplines. “You may not be changing minds, but you may be finding people

who have your back,” says Brooke Foucault Welles, a communications scientist at Northeastern University in Boston, Massachusetts. “And that's powerful.”

## BRAVE NEW WORLD

On 12 November 2014, the ESA Rosetta mission landed a spacecraft on a comet — a milestone for space exploration. But in certain corners of the Internet, Rosetta's landing day may be best remembered for the scantily clad women on Matt Taylor's shirt. Taylor, a Rosetta project scientist, sported the Hawaiian-style garment as he gave interviews to reporters at mission headquarters in Darmstadt, Germany, and answered questions on an ESA webcast. (His comments were also suggestive: Rosetta “is sexy, but I never said she was easy”, he told viewers.) It wasn't long before people following the historic comet landing took notice — and took to Twitter.

“What a lost opportunity to encourage girls into science,” tweeted Fernanda Foertter, a computer programmer at Oak Ridge National Laboratory in Tennessee. Others approached it with a bit more snark: “No no women are toooooooally welcome in our community, just ask the dude in the shirt,” wrote New York-based science journalist Rose Eveleth, who linked to a *Nature* video interview with Taylor.

What started as a trickle of tweets soon became a flood. By 14 November, the day that Taylor gave a tearful public apology on another ESA webcast, Twitter users had posted more than 3,100 messages using the #shirtstorm hashtag (see ‘Anatomy of a Twitterstorm’).

In many ways, #shirtstorm and other Twitter conversations about sexism are not new. It is only the venue that has changed, says Hope Jahren, a geobiologist at the University of Hawaii at Manoa who is active on Twitter. “Guys have been wearing girly shirts forever,” she says. “The women around them have been rolling their eyes and going home and saying, ‘What a buffoon. I'm so sick of this crap.’ They've been doing it in the women's room and doing it in the coffee room.” But now, Jahren says, “Twitter is that thought under your breath.”

The social-media service is also an enormous megaphone that claims to have 320 million active users each month. Research suggests that hashtag-driven Twitter conversations can help to amplify the voices of people who are not powerful by conventional measures.

One example comes from Foucault Welles and her colleagues' analysis of a hashtag that arose after police in Ferguson, Missouri, shot an unarmed African American teenager in August 2014. The killing quickly became a national news story, and the #ferguson hashtag became part of a broader US debate over police violence. Yet more than a year later, one of the most retweeted #ferguson contributors was a teenager from the Ferguson area.

“People who don't have power really can have their voices heard,” Foucault Welles says. “They can reframe the story.”

And that can make Twitter an important outlet for younger scientists, who often don't know how to respond to instances of sexism or sexual harassment. One 2014 survey of 666 scientists — including 516 women — found that 64% had experienced sexual harassment, and only 20% of that group said that they knew how to report such behaviour<sup>1</sup>. Most were students or postdoctoral researchers at the time they were harassed.

When scientists talk about sexism and harassment on Twitter, it presents younger researchers with a model for confronting such issues. “This way, they can see other people are going through it, and there is a positive effect to speaking out,” says Zuleyka Zevallos, a sociologist who manages the Science Australia Gender Equity Project at the Australian Academy of Science in Canberra.

For Ingleby, venting about her sexist journal review on Twitter paid unexpected dividends. She and her co-author, both postdocs, had waited three weeks for *PLoS ONE* to decide whether to grant their appeal and re-examine their paper. By making their plight public, Ingleby drew public support from other scientists — and, privately, invaluable advice from more-experienced researchers about how to deal with the journal. “I did get some messages that called me a feminazi and all that stuff,” Ingleby says, “but that was by far the minority.”

➔ **NATURE.COM**  
To hear a podcast discussion about Twitterstorms, go to: [go.nature.com/vkydzl](http://go.nature.com/vkydzl)

She has one crucial piece of advice for those who may follow in her footsteps: “Be a bit more prepared for things going viral. Maybe pick a few quiet days in your calendar.”

## IN THE EYE OF THE HURRICANE

Determining which factors can fan a handful of messages into an Internet firestorm, or what gives a hashtag staying power, is tricky. One study<sup>2</sup>, published in 2012 by researchers at the University of Pennsylvania in Philadelphia, suggests that Internet content goes viral when it elicits a strong emotional reaction. Marketing researcher Jonah Berger and decision scientist Katherine Milkman analysed the popularity of 6,956 news stories posted to the *New York Times* homepage between 30 August and 30 November 2008. The pair found that stories that inspired intense positive emotions, such as awe or amusement, were the most likely to go viral; anger, anxiety and other strong negative feelings also propelled articles to wide readership, but sadness seemed to reduce the chance that a reader would share a story with others. The recent science Twitterstorms, which are often fuelled by a combination of frustration, anger and black humour, fit with those ideas.

Yet an element of randomness is also at play. Joseph Reagle, a communications researcher at Northeastern University, sees this in the story of Cecil, a lion killed by an American tourist in Hwange National Park in Zimbabwe in July. The animal's death became an international cause célèbre, inspiring a hashtag (#CeciltheLion) that racked up 1.2 million tweets in one month — despite the fact that hunters kill dozens of lions in Zimbabwe each year<sup>3</sup>.

To Reagle, Cecil's tale also suggests that ‘hashtag activism’ is here to stay. “We are seeing the emergence of a genre,” he says. “And we will see it repeated.”

The conversations sparked by popular hashtags can shift the focus of media coverage and broader public discussion. The #YesAllWomen hashtag began in May 2014, in response to a shooting spree in California in which the killer said that his motivation was a hatred of women. Women used the hashtag to connect this violent misogyny to examples of everyday sexism and harassment — giving rise to a new wave of media coverage<sup>4</sup>. “That’s one of the really interesting things that starts to happen with some hashtags — they become news in their own right,” says Samantha Thrift, a feminist media scholar at the University of Calgary in Canada.

Hunt learned about the amplifying power of social media the hard way on 8 June. “You fall in love with them, they fall in love with you, and when you criticize them, they cry,” he said in a speech at the World Conference of Science Journalists in Seoul. His comments were tweeted by audience members, creating an Internet furor that quickly hit mainstream news outlets.

On 10 June, Hunt told BBC Radio 4 that he was “really sorry”. But in later comments to *The Observer*, he said that he had been “hung out to dry” and forced to resign an honorary post at University College London. “It has done me lasting damage,” he said. “What they did was unacceptable.”

For those in positions of power, such as Hunt, finding themselves at the centre of a Twitterstorm can be deeply unsettling, given social media's ability to upend traditional hierarchies. But many women who talk about sexism, feminism and gender issues online face a harsher reception, from abusive comments to threats of physical harm. When Eveleth tweeted her criticism of Taylor's shirt, she received death threats. When others joined the fray, such as Jacquelyn Gill, a palaeoecologist at the University of Maine in Orono, they became targets, too.

“I stand with @roseveleth and others who are calling out sexism despite online harassment,” she tweeted. “I’m reporting abusive tweets as I’m able.” She added: “Free-speech apparently only applies to dudes

threatening violence to women with an opinion — not the women with an opinion. #shirtstorm”.

The reaction to her commentary was swift and punishing. “For the next 72 hours I got death and rape threats,” Gill says. “It was a non-stop barrage of people trolling those hashtags.”

As the stream of vitriol became overwhelming, some of Gill's colleagues wrote a computer program to scan Twitter for threatening messages that mentioned her username. That spared Gill from constantly monitoring her account for serious threats. But no program could spare her from the

awkward conversations that she had with University of Maine officials after realizing that some of her harassers on Twitter were discussing how to get her fired in retaliation for her ‘Shirtgate’ activism.

“I’ve run up against the real-world consequences of speaking as a woman on the Internet,” she says.

This problem is not limited to science: in a study of 2,849 Internet users, the Pew Research Center in Washington DC reported that 40% had been harassed online. Although men are more likely to be called offensive names or purposefully embarrassed, women are more likely to be stalked or sexually harassed as a result of their Internet use. The survey also found that social media is the place where women are most vulnerable to harassment of all types, ranging from stalking to physical threats.

Faced with such attacks, some scientists have begun to rethink how they participate in online discussions about sexism. Some retreat entirely; others, wary of being silenced by abuse, try to find safer ways to engage online. One female researcher who has suffered Internet harassment now tweets about feminist issues under a pseudonym while also maintaining an active Twitter account under her real name. “It makes me feel safer,” says the researcher, who asked not to be named. “Although, in a lot of these cases, if someone wants to find you, they will.”

## MAKING SENSE OF A MOVING TARGET

Researchers tracking the rise of social media are trying to understand whether intense discussions online translate into real-world change. The difficulty lies in deciding how to measure such effects.

One approach draws on network analysis. A team of computer scientists at Carnegie Mellon University in Pittsburgh, Pennsylvania, tracked Twitter users' interactions before, during and after 20 Twitterstorms between 2011 and 2014 — most centred on targets of broad interest, such as US late-night television host Stephen Colbert and fast-food chain McDonald's<sup>5</sup>. The researchers found that these events did not create lasting links between participants, as measured by who these users follow or message on Twitter. This suggests that Internet dust-ups do not usually lead to sustained discussion or greater awareness of a given issue.

But other studies show that intense Twitter discussions may affect contributors in ways that are harder to quantify. Mindi Foster, a social psychologist at Wilfrid Laurier University in Waterloo, Canada, decided to investigate the psychological effects of tweeting on the basis of her own experience using social media. After hearing an anti-Semitic remark on a television programme one night, Foster joined Twitter to vent her anger — and it felt good.

Foster's research seems to confirm her hunch: that when women tweet about sexism, it improves their sense of well-being<sup>6</sup>. The study involved 93 female university students who were presented with information about sexism in academia, politics and the media. One group of students was asked to tweet publicly about what they had learned, another to tweet privately and a third to tweet about the weather. (A fourth group was told to do nothing.)

During the three-day study, each participant filled out a daily questionnaire on her emotional state. Those who were assigned to tweet publicly reported a greater sense of well-being, on average, by the end of the experiment; those in the other groups showed no change. These results,

# “I’VE RUN UP AGAINST THE REAL-WORLD CONSEQUENCES OF SPEAKING AS A WOMAN ON THE INTERNET.”



# ANATOMY OF A TWITTERSTORM



The shirt

On 12 November 2014, the European Space Agency (ESA) landed a spacecraft on a comet as part of the Rosetta mission. On the same day, ESA scientist Matt Taylor wore a shirt covered with half-naked women to mission control, prompting Twitter discussions under the #shirtstorm and #Shirtgate hashtags about the message that the garment sent to women and girls. Two days later, palaeoecologist Jacquelyn Gill started the #scishirt hashtag to encourage scientists to post pictures of their favourite (non-sexist) shirts.

## HASHTAG DEMOGRAPHICS

Tens of thousands of tweets about the shirt may have been seen by tens of millions of people (reach). Gender estimates (Twitter does not ask for this information) suggest that the majority of tweets using #shirtstorm and #Shirtgate came from men.

**#shirtstorm**  
53,905 tweets  
Gender split:  
22% female

**#Shirtgate**  
38,457 tweets  
Gender split:  
21% female

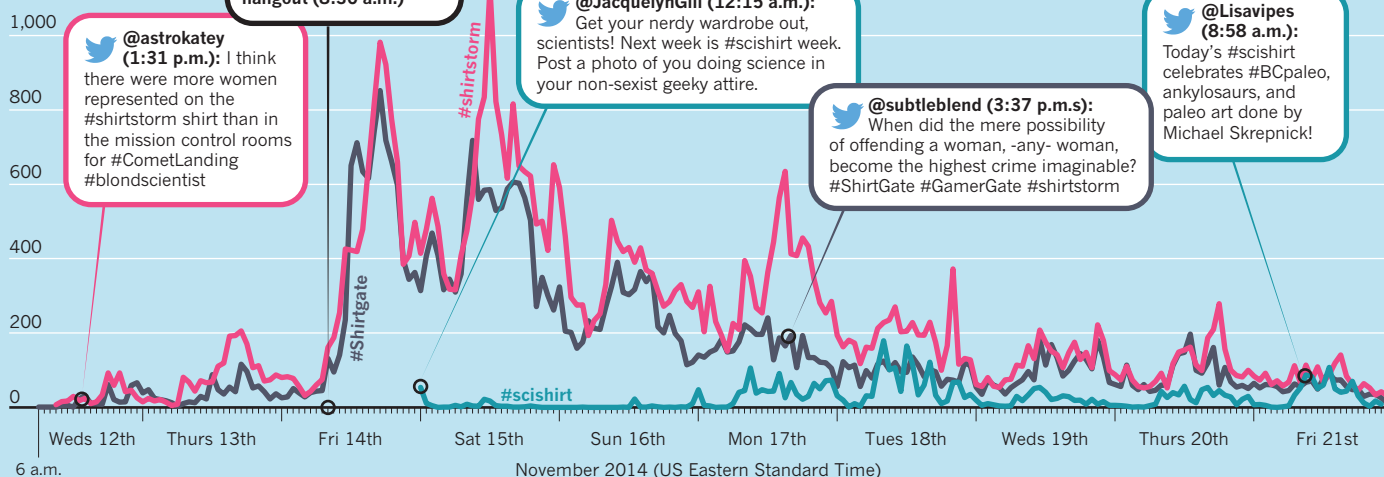
**#scishirt**  
4,424 tweets  
Gender split:  
67% female

Reach:  
49.5 million  
unique users

Reach:  
32.7 million  
unique users

Reach:  
5.5 million  
unique users

Number  
of tweets  
1,200



SOURCE: KEYHOLE; SHIRT IMAGE: NATURE VIDEO

although preliminary, are in line with earlier research that shows that expressive writing — such as writing in a diary — can provide similar benefits. But Foster speculates that public tweeting may confer an extra boost because it spurs writers to think more deeply about what they are saying.

Twitter can also help to build a sense of community among scientists in different disciplines who are confronting sexism and sexual harassment. Sometimes these bonds grow out of dark humour, such as the #distractinglysexy hashtag birthed in reaction to Hunt's comments. Thousands of female researchers posted pictures of themselves in labs and at field sites, up to their knees in mud or swathed in shapeless biosafety suits. "Filter mask protects me from hazardous chemicals and muffles my woman cries," wrote Amelia Cervera, a biochemist at the University of Valencia in Spain, who shared a photo of herself wearing the face-obscuring gear.

Gill, a palaeoecologist, says that she has begun to connect with researchers in astronomy, anthropology, engineering and computer science, among other fields. Such links can help researchers to learn from each other's experiences of confronting sexism. "Some of our disciplines have been better at gender equality than others," she notes. "Some of us have been having these discussions for a long time."

But the ongoing Twitter conversation about sexism is also limited in some important ways. It often ignores the concerns of women whose experiences with sexism are exacerbated by discrimination on the basis of race, sexual orientation or disability. For example, a US survey of 557 female scientists from ethnic minority groups found that two-thirds felt pressure to prove themselves over and over again — beyond what was asked of white colleagues. And 48% of African American respondents said that they had been mistaken for janitors (caretakers) or administrative staff in their workplaces.

"If you are a minority within a minority, you are actually dealing with multiple problems," says Zevallos. That is just as true on Twitter as it is in the lab or office.

And this can make women who are dealing with the effects of multiple

forms of discrimination feel excluded from conversations that focus on sexism or sexual harassment. Such concerns surfaced recently in the wake of the Marcy sexual-harassment case, which had prompted a vigorous online debate under the #astroSH hashtag. "If you are not talking about and confronting racism with same vigilance as sexism, might as well hang 'no Blacks' signs," tweeted Chanda Prescod-Weinstein, an astrophysicist at the Massachusetts Institute of Technology (MIT) in Cambridge. "And I say that as a victim of both sexual assault and sexual harassment." Sarah Ballard, also an MIT astrophysicist, echoed the sentiment: "We can't rely on crowdsourcing meting out of justice- (Mostly white) crowds will stand up for white women, \*crickets\* otherwise."

And although social media can help to create a community discussion about sexism and other forms of discrimination, fighting for equality requires the real-world cooperation of universities, governments and other institutions. Some of these have taken action in response to sexist incidents that online discussions helped to bring to wider attention. But although Twitter may be hard to ignore, it does not have the authority to set and enforce expectations for fair treatment.

Despite those caveats, Thrift finds great value in the ongoing social-media conversations among scientists, which she sees as a form of public education — and the first step towards concrete change. "That's hugely important," she says. "If we don't name something as sexist, as harassment, as misogyny, it will continue unchecked." ■

**Lauren Morello** is a senior news editor for Nature in Washington DC.

1. Clancy, K. B. H., Nelson, R. G., Rutherford, J. N. & Hinde, K. *PLoS ONE* **9**, e102172 (2014).
2. Berger, J. & Milkman, K. L. *J. Marketing Res.* **49**, 192–205 (2012).
3. Lindsey, P. A. et al. *PLoS ONE* **8**, e73808 (2013).
4. Thrift, S. C. *Feminist Media Stud.* **14**, 1090–1092 (2014).
5. Lamba, H., Malik, M. M. & Pfeffer, J. A tempest in a teacup? Analyzing firestorms on Twitter. In *Proc. 2015 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining* (in the press).
6. Foster, M. D. *Br. J. Soc. Psychol.* <http://dx.doi.org/10.1111/bjso.12101> (2015).



# NEW LIFE FOR PIG ORGANS

*Gene-editing technologies have breathed life into the languishing field of xenotransplantation.*

BY SARA REARDON

**P**ale on its bed of crushed ice, the lung looks like offal from a butcher's counter. Just six hours ago, surgeons at the University of Maryland's medical school in Baltimore removed it from a hefty adult pig and, with any luck, it will soon be coaxing back to life, turning a rich red and resuming its work in the chest of a six-year-old baboon.

An assistant brings the lung to Lars Burdorf and his fellow surgeons, who currently have their hands in the baboon's splayed chest. The team then begins the painstaking process of connecting the organ to the baboon's windpipe and stitching together the appropriate arteries and blood vessels. But this 5-hour, US\$50,000 operation is just one data point in a much longer

experiment — one that involves dozens of labs and decades of immunological research and genetic engineering to produce a steady and safe source of organs for human transplantation. If the baboon's immune system tolerates this replacement lung, it will be a sign that the team is on the right track.

Robin Pierson heads the Maryland lab, which has performed about 50 pig-to-primate transplants like this one to test different combinations of genetic modifications in the pig and immune-suppressing drugs in the primate. Even so, the team has not had a primate survive for longer than a few days. The complexities of the immune system and the possibility of infection by pig viruses are formidable and

drove large companies out of the field in the early 2000s.

That trend may now be reversing, thanks to improved immunosuppressant drugs and advances in genome-editing technologies such as CRISPR/Cas9. These techniques allow scientists to edit pig genes, which could cause rejection or infection, much more quickly and accurately than has been possible in the past. In October, eGenesis, a life-sciences company in Boston, Massachusetts, announced that it had edited the pig genome in 62 places at once.

Some researchers now expect to see human trials with solid organs such as kidneys from genetically modified pigs within the next few years (see 'Choice cuts'). United Therapeutics,

CHRIS MADDALONI/NATURE



**Surgeons prepare a genetically modified pig lung for transplantation experiments.** a biotechnology company in Silver Spring, Maryland, has spent \$100 million in the past year to speed up the process of making transgenic pigs for lung transplants — the first major industry investment in more than a decade. It says that it wants pig lungs in clinical trials by 2020. But others think that the timeline is unrealistic, not least because regulators are uneasy about safety and the risk of pig organs transmitting diseases to immunosuppressed humans.

“I think we’re getting closer, in terms of science,” says transplant surgeon Jeremy Chapman of the University of Sydney’s Westmead Hospital in Australia. “But I’m not yet convinced we’ve surpassed all the critical issues that are ahead of us. Xenotransplantation has had a long enduring reality that every time we knock down a barrier, there’s another one just a few steps on.”

## LONG HISTORY

Surgeons have been attempting to put baboon and chimpanzee kidneys into humans since at least the 1960s. They had little success — patients died within a few months, usually because the immune system attacked and rejected the organ. But the idea of xenotransplantation persisted. It could, proponents say, help to save the lives of the tens of thousands of people around the world who die each year while waiting for a suitable human donor. And having a steady supply of farm-grown organs would allow doctors to place recipients on immunosuppressant drugs days ahead of surgery, which should improve survival rates.

When details about why non-human organs are rejected began to emerge in the 1990s, the transplantation field was ready to listen. In 1993, surgeon David Cooper of the University of Pittsburgh in Pennsylvania and his colleagues discovered that most of the human immune reaction was directed at a single pig antigen: a sugar molecule called  $\alpha$ -1,3-galactose, or  $\alpha$ -gal, on cell surfaces that can cause organ rejection within minutes<sup>1</sup>. An enzyme called  $\alpha$ -1,3-galactosyltransferase is necessary for producing this sugar, and knocking out the gene that produces the enzyme should temper the reaction.

This discovery and other advances in transplantation medicine made the problem seem more tractable to big pharmaceutical companies. In 1996, Novartis in Basel, Switzerland, began to invest heavily in xenotransplantation research, says Geoffrey MacKay, who was the firm’s business director for transplants and immunology at the time and oversaw the xenotransplantation effort. “They wanted to not only put a dent into the organ shortage but really solve it via transgenic pigs,” MacKay is currently interim chief executive at eGenesis.

Novartis initially planned to spend more than \$1 billion on xenotransplantation, including both scientific research and planning the infrastructure that would be needed to grow

pigs in germ-free facilities around the world. Other companies put some skin in the game, including Boston-based Genzyme and PPL Therapeutics, the British company that collaborated in the creation of Dolly, the first cloned sheep. Regulators such as the US Food and Drug Administration (FDA) began to draw up the guidance and standards that companies would need to meet before the technology could be moved into people.

But the immune system turned out to be much more complex than anticipated, and baboons that received pig organs never survived longer than a few weeks, even when the researchers were able to suppress  $\alpha$ -gal production with drugs. A second major concern, especially to regulators, was the risk of infection. Even if pigs could be kept entirely sterile, the pig genome is sprinkled with dozens of dormant porcine endogenous retroviruses (PERVs), and

## “I THINK PEOPLE ARE COMING AROUND TO LOOK AT XENOTRANSPLANTATION IN A MORE-FAVOURABLE LIGHT.”

studies conflicted as to whether these could become active in humans.

The challenges proved too daunting, and in the early 2000s Novartis killed its xenotransplantation programme, reshuffling or laying off its researchers. Other companies followed suit. It became, Pierson says, “the third rail of biotech to discuss xenotransplantation as a business plan”.

For the next ten years, the business side of the field went dark, at least as far as solid-organ transplants were concerned. Meanwhile, a few research teams and start-up companies began pursuing pig tissue transplants: a much simpler goal than using solid organs because the immune response is not as severe. In April, Chinese regulators approved the use of pig corneas from which all the cells have been removed<sup>2</sup>. Also on the near horizon are pig insulin-producing islet cells that might be transplanted into people with diabetes.

The first commercially available islets are likely to come from technology designed by Living Cell Technologies (LCT), a biotech company based in Auckland, New Zealand, that has developed a process to encapsulate pig islet cells in a gelatinous ‘dewdrop’ that protects them from the human immune system. The product, called DIABECCELL, is currently in late-stage clinical trials in several countries. Patients implanted with the cells have survived more than nine years without evidence of immune rejection or infection<sup>3</sup>.

“I think people are coming around to look at xenotransplantation in a more-favourable light knowing that we have strong safety data,” says LCT research lead Jackie Lee. Diatrans Otsuka Limited, in Auckland, is now running the programme.

Solid organs still pose a challenge. The handful of researchers who have continued to work with them have solved some of the problems that vexed Novartis, such as identifying other key pig antigens and the correct combinations of immunosuppressant drugs. But different organs have different problems: kidneys may be safer than hearts, for instance. Lungs, as Pierson’s team has discovered, are extremely difficult to transplant, because they have extensive networks of blood vessels, which provides more opportunities for primate blood to meet pig proteins and to coagulate. Pierson’s current trials use lungs from an  $\alpha$ -gal-knockout pig that includes five human genes. The baboon is treated with a combination of four immunosuppressant drugs.

Most US researchers, including Pierson and Cooper, have relied on pigs made by a regenerative-medicine company called Revivacor in

Blacksburg, Virginia, that spun-out from PPL Therapeutics. In 2003, Revivacor co-founder David Ayares and his colleagues created the first cloned pig genetically modified to delete  $\alpha$ -gal<sup>4</sup>. The company has since been experimenting with altering other protein antigens that trigger the immune system or cause human blood to coagulate.

These modifications have greatly lengthened the time that an organ can survive in a baboon. In one trial, surgeon Muhammad Mohiuddin at the National Heart, Lung, and Blood Institute in Bethesda, Maryland, and his colleagues took the heart from an  $\alpha$ -gal-free pig that had two human genes that protect from coagulation and sewed it into the abdomen of a baboon<sup>5</sup>. The organ did not replace the baboon’s heart, but the animal lived with the implant for two and a half years.

Mohiuddin says that the group is now attempting a ‘life-supporting’ transplant by replacing the baboon’s heart with a pig heart. The longest life-supporting transplant was published in June<sup>6</sup>, when Cooper’s group announced that a kidney transplant from a Revivacor pig with six modified genes supported a baboon for 136 days.

## GENERATION GAME

But the process is slow, Cooper says. It generally takes several generations of breeding to knock out both copies of just one given gene in a pig. Deleting multiple genes or swapping them for their human counterparts takes many more generations, because every litter contains pigs with different combinations of the modified genes.

That is why so many are excited about precise genome-editing tools such as

CRISPR/Cas9, which can precisely cut both copies of a gene — or genes — straight from a pig embryo in one go. “Our first [ $\alpha$ ] gal-knockout pig took three full years,” says Joseph Tector, a transplant surgeon at Indiana University in Indianapolis. “Now we can make a new pig from scratch in 150 days.” His group recently used CRISPR to knock out two pig genes simultaneously<sup>7</sup>. The researchers are now beginning to transplant CRISPR-modified pig organs into macaques, one of which has survived for more than three months.

Eventually, gene editing might even eliminate the need for immunosuppression, says Bernhard Hering, a transplant surgeon at the University of Minnesota in Minneapolis. His group is using CRISPR to create pig islets that could be transplanted without the need for drugs. Partly because of LCT’s success with encapsulated islets, many are hopeful that islet cells will be the first genetically modified tissue to make it into clinical trials, paving the regulatory pathway for the more-difficult organs. A non-profit organization has built a germ-free facility in which to raise Hering’s pigs.

## TECHNOLOGY REVIVAL

The gene-editing advances have brought new investment into the field. In 2011, United Therapeutics acquired Revivicor for about \$8 million and announced an ambitious plan to start clinical trials of gene-edited pig lungs by the end of the decade. The company’s co-chief executive, Martine Rothblatt, secured land in North Carolina for a farm that could produce 1,000 pig organs per year and says she expects to break ground by 2017. The facility’s elaborate plans include solar panels and helicopter landing pads to help speed fresh organs to those in need.

In 2014, United Therapeutics formed a \$50-million partnership with the biotech firm Synthetic Genomics (SGI) in La Jolla, California, founded by genome-sequencing pioneer Craig Venter. Rather than simply knocking out antigens, SGI is also engineering tissues that sidestep rejection in a different way — such as pig cells that produce surface receptors that act as ‘molecular sponges’ and sop up human immune signalling factors that would otherwise attack the organ. CRISPR and other methods also allow the researchers to make tweaks such as lowering a gene’s expression rather than deleting it completely, says Sean Stevens, head of SGI’s mammalian synthetic-biology group. In September, United Therapeutics committed another \$50 million.

Peter Cowan, an immunologist at St Vincent’s Hospital in Melbourne, Australia, is taking a different approach. His group has made pigs that generate antibodies against human immune cells. In their design, the antibodies would be made only by transplanted liver cells, ensuring that the immune system is suppressed just around the organ.

eGenesis was founded in April by bioengineer

## CHOICE CUTS

Researchers are looking to source an increasing variety of living tissues, including solid organs, from pigs. Many are attempting to genetically engineer the animals to reduce the risk of rejection and infection in humans.

### CORNEA

Pig corneas were approved for marketing in China in April.

### HEART

A genetically modified pig heart implanted in a baboon’s abdomen survived for 2.5 years.

### LUNG

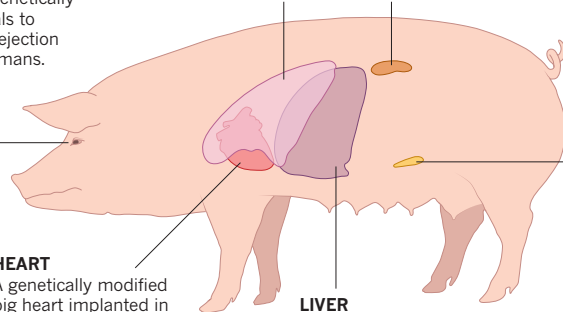
A factory farm is being designed to produce 1,000 pig lungs per year.

### KIDNEY

A kidney with six genetic modifications supported a baboon’s life for 4 months.

### PANCREAS

Phase III clinical trials of insulin-producing islet cells are under way.



Luhan Yang and geneticist George Church of the Wyss Institute and Harvard University in Cambridge, Massachusetts. MacKay says that the firm plans to begin transplanting organs into primates next year. To that end, Church says that the company has made embryos that have more than 20 genetic alterations to cell-surface antigens and other factors and is ready to implant the embryos into female pigs. One of its first publications used CRISPR to inactivate 62 occurrences of PERV genes in pig kidney cells<sup>8</sup>. The researchers have since transferred the cells’ nuclei into pig embryos.

Incidentally, few researchers in the field see the PERV problem as a major safety concern. The virus replicates poorly in human tissues and the risk of spreading it is virtually non-existent, says Jay Fishman, an infectious-disease specialist at Massachusetts General Hospital in Boston. He says that researchers have tracked dozens of people who received unregulated porcine skin grafts, and none seems to have developed disease.

But dealing with PERVs may be a regulatory necessity. The FDA said in an e-mail to *Nature* that it is still concerned about the possibility of disease caused by PERVs. There are other pathogens to worry about, too. Most major epidemics start with an animal pathogen that jumps to humans, warns Peter Collignon, an infectious-disease scientist at the Australian National University in Canberra. “If you want to do the perfect experiment for finding new novel viruses and letting them multiply, this is it.”

Unless xenotransplants are proved to be extremely safe, the FDA suggests that they be limited to people with life-threatening conditions who have no other options. It will be even harder to get organs from genetically modified pigs to market, the agency says, because regulators must approve both the genetic construct used to make the animal and the organ itself.

Even if safety can be assured, questions remain about whether pig organs would work correctly in their new home, Chapman says.

It is unclear whether a pig kidney would, for instance, respond to the human hormones that regulate urination, or whether proteins produced by a pig liver would interact correctly with human systems. And because pigs live for only about ten years, their organs might not survive a human lifetime. Even using a xenotransplant as a ‘bridge’ until a suitable human donor is found will be difficult. After a heart transplant, for instance, fibrous tissue forms around the new organ, making second transplants very difficult, Chapman says.

Given the long list of known hurdles, the surprise setbacks that researchers encounter along the way can be particularly disheartening. About half an hour after its surgery at the University of Maryland, the baboon with a pig’s lung woke up in a cage wearing a small vest that monitored its vital signs. The lung functioned well overnight and was even able to provide enough oxygen to the animal when blood flow to its other lung was temporarily blocked. But the next day, the animal became ill and had to be killed. That was unexpected, Pierson says, because the pig’s multiple genetic modifications seem to have worked well with the baboon’s immune system. A post-mortem examination revealed that fluid had accumulated in the lung and the organ had developed blood clots. Like so many other aspects of xenotransplantation, Pierson says, “this is a problem that we are still learning about.” ■

Sara Reardon writes for *Nature* in Washington DC.

1. Oriol, R., Ye, Y., Koren, E. & Cooper, D. K. *Transplantation* **56**, 1433–1442 (1993).
2. Zhang, M. C. et al. *Am. J. Transplant.* **15**, 1068–1075 (2015).
3. Elliott, R. B. et al. *Xenotransplantation* **14**, 157–161 (2007).
4. Phelps, C. J. et al. *Science* **299**, 411–414 (2003).
5. Mohiuddin, M. M. et al. *J. Thorac. Cardiovasc. Surg.* **148**, 1106–1113 (2014).
6. Iwase, H. et al. *Xenotransplantation* **22**, 302–309 (2015).
7. Higginbotham, L. et al. *Xenotransplantation* **22**, 221–230 (2015).
8. Yang, L. et al. *Science* <http://doi.org/83q> (2015).



# COMMENT

**BIOETHICS** Four questions face delegates of gene-editing summit **p.159**

**CANCER** A stirring memoir of crude therapies and internecine politics **p.162**

**HISTORY** Kepler cast women as knowledge-makers to save his mother **p.164**

**EMISSIONS** Russia's coast holds rich potential for renewable-energy generation **p.165**



DOUGLAS C. PIZAC/AP/PA IMAGES



Biosafety-level-3 protection at the US Army's Dugway Proving Ground, Utah.

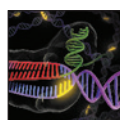
## Rethink biosafety

**Tim Trevan** calls on those working with organisms that are hazardous, or could be so, to take lessons from the nuclear industries, hospitals and other sectors that have established a safety culture.

**T**wo months ago, the US Department of Defense froze operations at nine biodefence laboratories where work is done on dangerous pathogens. Inspectors had discovered live anthrax outside a containment area at the US Army's Dugway Proving Ground — a facility in Utah that tests defence systems against biological and chemical weapons.

The discovery at Dugway is the latest of several concerning finds. In June 2014,

workers at a US Centers for Disease Control and Prevention (CDC) biosafety-level-3 laboratory in Atlanta, Georgia, sent anthrax samples to three other laboratories on the same campus. The samples were meant to have been sterilized but several



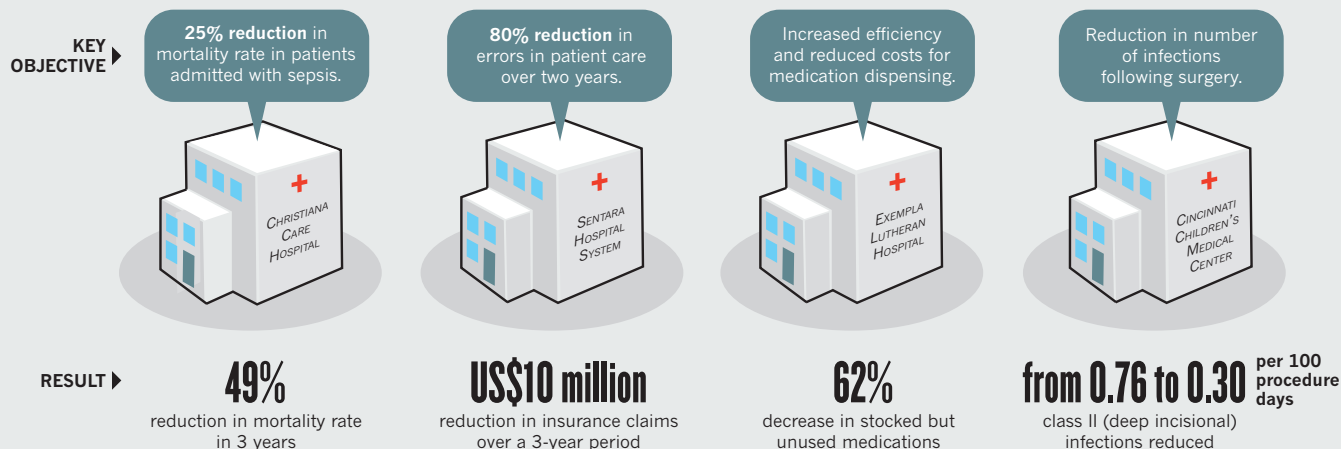
**CRISPR GENE EDITING**  
A Nature collection  
[nature.com/crispr](http://nature.com/crispr)

factors meant that 41 people were potentially exposed to live bacteria<sup>1</sup>. Then in May this year, an investigation revealed that for several years, staff at Dugway had been improperly sterilizing anthrax samples, and that live spores may have been sent to 52 laboratories in the United States, Canada, Australia and South Korea.

These mishaps — which are by no means unique to anthrax — are worrying on two levels. First, the handling of ►

## FOLLOW THE LEADER

Hospitals throughout the United States are taking steps to become 'high-reliability organizations' (HROs). The US Centers for Disease Control and Prevention (CDC) should follow suit.



SOURCES: M. T. ZUBROW ET AL. *JT COMM J QUAL PATIENT SAF* 34, 187–191 (2008); AHRQ, USDHHS

► dangerous pathogens within a controlled environment is one of the easier biological risks to contain. Much harder is ensuring that basic biological research that is known to be potentially dangerous, or that turns out to be so, is carried out safely. Second, it is only going to get harder to ensure the safe and secure use of organisms and their products — whether in basic research or in detecting and preventing the development of biological weapons.

Relatively inexpensive and easy-to-use tools and approaches are greatly expanding the possibilities for genetic engineering, including for would-be terrorists. Among them are the gene-editing technique CRISPR/Cas9, and the use of gene drives — where the biased inheritance of particular genes alters entire populations. Meanwhile, myriad developments are undermining existing approaches to non-proliferation. These include: the sale of equipment and materials over the Internet; the accessibility of computing power; and the rise of the open-science movement.

What are the prospects for managing the more intractable risks globally if measures to ensure the safe handling of dangerous pathogens are failing at the best-equipped facilities in the country with the most advanced biotechnology in the world? The anthrax incidents occurred despite the use of extensive legislation, protocols and procedures. The problem with the CDC, the US Department of Defense, and the many labs around the world who follow their lead, is not a lack of knowledge or training, or even a lack of engineering resources. It is the lack of a safety culture.

Most laboratories handling potentially dangerous biological materials are stuck in compliance mode. To prevent human and environmental catastrophes, and the

shutdown of important research, that mindset must be transformed.

I never thought I'd write this, but I believe that it is time for experts who advise on biosafety and biosecurity to learn from specialists in nuclear security. I define biosafety and biosecurity as the prevention of the accidental release of potentially harmful organisms or their products and the prevention of the deliberate release of such agents for nefarious purposes. Leaders in these areas include the CDC, the World Health Organization (WHO), and Public Health England in the United Kingdom.

### OUTSIDE THE BOX

The reluctance of those of us in biosecurity and biosafety to learn from the nuclear industry stems from the fact that many of the practices in nuclear security and safety are not transferable to biology. For instance, monitoring the amount of materials entering and leaving a complex makes little sense when a tiny sample can contain militarily significant amounts of a hazardous substance. And expensive security measures — guns, gates, guards and cameras — make sense at nuclear power plants, of which there are only a few hundred worldwide. They are not feasible at the much greater number of labs and hospitals dealing with hazardous biological agents. Moreover, hospital accident-and-emergency buildings and procedures are designed to get patients inside as quickly as possible, not keep them out. And progress within public health and research depends on transparency and open collaboration.

What those working with biologicals can learn from practitioners in the nuclear industry — as well as from those in the US Navy, offshore oil drilling, airlines and utilities — is a culture of safety. In all these

areas, best practice focuses on preventing failure rather than on maximizing output. The result is what is called a 'high-reliability organization' (HRO).

HROs feature the following five characteristics<sup>2</sup>. First, everyone within the organization constantly asks, 'What can go wrong and how do we prevent it?' Second, workers are sensitive to any deviation from the norm, such as an unexpected change in the temperature of the reactor core in the case of a nuclear power plant, and learn to ascertain which variances can snowball into catastrophic failure. Third, systems are designed to be resilient so that if they do fail, they do so with minimal damage and recovery can be quick. Fourth, workers recognize that the operating environment is complex and changeable, and that mindlessly following standard procedures without paying attention to what else is going on in the environment can be dangerous. Lastly, expertise is valued over seniority, with the recognition that it may be the newest or most junior member of a team who spots a problem or who knows best how to fix it.

In HROs, safety is not 'for them' but 'for each and every one of us', and is seen as an investment rather than a short-term cost. Workers are encouraged to hold each other accountable and to report red flags, such as a change in behaviour that might make a colleague more prone to mistakes. Mishaps and near misses too are reported without fear of blame, and mistakes are analysed to learn how to prevent them from recurring. Finally, the process is one of continual improvement: attention to safety does not stop just because certain targets have been met.

In biosecurity and biosafety, the CDC is widely seen as the global gold standard. The CDC's handbook *Biosafety in Microbiological and Biomedical Laboratories* has become



the reference for laboratories worldwide. Other resources that it provides (posters, training videos, data and information) along with documents from the WHO, are used as core reference materials, even in the most remote labs. Yet the world's exemplars in the handling of the most dangerous pathogens, and therefore the multitude of public and private organizations who follow them, are stuck in a very different culture from that of HROs.

From the CDC to diagnostics and basic-research laboratories worldwide, the emphasis is on ticking boxes and on following rules set by outside authorities, such as the Department of Health and Human Services in the United States or the relevant agencies in the European Union. Safety is generally seen as an inconvenience that detracts from the main task at hand. It is delegated to biosafety officers, and after-the-fact indicators of problems such as the number of accidents, are the predominant metric, with the implicit aim being to ensure that spills, infections and so on are kept below targets with minimal effort.

A recent illustration of problems caused by the rote following of rules is the handling of an Ebola patient by staff at Texas Health Presbyterian Hospital in Dallas in 2014, where two nurses contracted the disease. Having never dealt with a suspected Ebola case before, staff checked the CDC website for information on the correct personal protective equipment (PPE) to wear. Unfortunately, that website described PPE

more suitable to handling Ebola samples in a laboratory. The PPE the hospital workers initially used left areas of their face and necks exposed.

Failure to consider context and all the links in the chain can similarly undermine the value of spending millions of dollars on building and operating containment labs throughout the world. A recent inspection at a major diagnostic lab for animal diseases in Afghanistan, for instance, revealed that standard operating procedures (SOPs) copied from Western labs, for 'safe' operation, were being followed to the letter, including one for the handling of biological waste. The waste was getting bagged up pending incineration. But because there was no budget for petrol for the incinerator, the bags were simply being stored, undermining many of the prior biosafety procedures.

### SAFETY FIRST

Organizations that have successfully implemented a culture of safety have often treated the introduction of a new way of doing things as a business project, akin to a move to a new software platform. Experts in the offshore oil industry have likened the process to moving from directing one play to another<sup>3</sup>. One must deal with a new script (the vision), a new stage set and scenery (the facilities, equipment and technology used in operations), new stage directions (processes and procedures), new roles (job descriptions), new contracts (hires), and new

rehearsals (training and commissioning).

For example, the metals manufacturer Alcoa, based in New York, launched a safety drive starting in the late 1980s using such techniques and saw the average rate of lost workdays (due to work-related injuries) drop over a ten-year period from 1.86 to just 0.18 per 100 work years<sup>4</sup>. As well as this willingness to start afresh, three other steps are crucial.

**Provide leadership, funds, time and commitment.** The process starts with senior

*"People who refuse to adapt should lose their positions."*

management laying out what safety means for their particular organization. All layers of the organization are then involved

in identifying what facilities, equipment and practices need to be changed. Lastly, a master plan is drawn up to realize the vision.

In some cases, considerable sums will be needed initially. Yet such investments can quickly pay off. Alcoa, for instance, jump-started its safety programme by spending US\$3 million over two months to fix unlit passageways in its plants. But based on the US Department of Labor's Accident Cost Calculator, the reduction in time lost due to accidents saved Alcoa around \$51.5 million annually.

**Make safety matter to everyone.** People will care about safety at their organization if their immediate bosses and those at the top frequently talk about it and back their talk with actions. If other achievements, such as efficiency or the output of journal papers, are rewarded ahead of safety — as is the case in most basic-research labs — people will pay less heed to it.

Those who ignore new safety rules must face significant sanctions. People who refuse to adapt should lose their positions. Various tools can aid managers on this front. For instance, workers can be required to obtain certification before being allowed to perform potentially hazardous tasks.

**Exploit peer accountability.** Most managers of staff who have employment protection, such as tenured professors or civil servants, cannot hire and fire, or give or withhold bonuses. Fortunately, cash does not seem to be a key motivator when it comes to safety.

A 2010 study of 1,600 safety professionals across different industries found that people's expectations of their peers seems to be the most important influence on workplace behaviour — ahead even of management's expectations<sup>5</sup>. And recognition of a job well done can be more motivating than a bonus. For instance, the Gallup Organization, based

JIM SUGAR/CORBIS



Workplace injuries plummeted when metals manufacturer Alcoa overhauled its approach to safety.





People working in the nuclear industry are encouraged to ask 'What can go wrong?'

in Washington DC, has surveyed more than 4 million workers worldwide, and found that employees who are recognized for their achievements have better safety records and fewer accidents on the job.

#### LEVERAGE LEADERSHIP

In 2011, the Joint Commission, a non-profit organization that controls hospital certification in the United States, started promoting HRO approaches in hospitals throughout the country<sup>6</sup> (see 'Follow the leader'). This followed several serious medical errors, such as surgeons operating on the wrong side of the brain in three patients in one year at the Rhode Island Hospital in Providence. In the case of the hospitals, the actual procedural changes — anything from more-stringent processes for infection control to improved systems for record checking — vary from place to place, but the aim is always to minimize the chances of something going wrong.

The CDC is the obvious candidate to pick up the torch and prove that the HRO approach also works in laboratory settings. It has the resources. And where the

*"The CDC is the obvious candidate to pick up the torch."*

CDC leads, others follow; if they do not, they risk not being able to acquire funding, collaborate with those in other laboratories or obtain contracts from corporations who demand compliance with best practice.

Changing the culture of such a large entity will be difficult. But proof of concept could be achieved first in one unit, such as the Bioterrorism Rapid Response and Advanced Technology Laboratory (BRRAT). Because BRRAT is on the CDC's main campus in Atlanta, top managers from across the organization could more easily be engaged in the process of organizational change. Approaches used at BRRAT could then be rolled out to the entire organization.

The HRO approach will be especially valuable for those facing uncertainty. In experimenting with gene drives and CRISPR/Cas9, there are no SOPs to follow. Asking 'What could go wrong?' or 'How could this science be misused?' and 'How can we prevent that from happening?' will embed biosafety and biosecurity considerations into study programmes from the outset.

Although research will always have an element of the unknown, under the HRO model, workers are encouraged to constantly monitor outcomes against expectations and

to make adjustments on the basis of new evidence. In other words, an HRO approach means reappraising biosafety and biosecurity plans as understanding increases.

The biological-research community is capable of taking this road: people working on gene drives, for instance, are actively debating potential safety and security issues<sup>7</sup>. But HRO principles need to be adopted much more widely.

Failure to do so could greatly harm society, agriculture and the environment. Take, for instance, the 2007 release of foot-and-mouth disease virus from the Pirbright Institute, an animal-health research centre in Woking, UK. Inadequate sterilization of biological waste, broken waste pipes and unsealed and overflowing manhole covers led to more than 2,000 sheep and cows being slaughtered, at a cost of \$200 million.

Moreover, failure to be seen to be conducting biology in a safe, responsible and ethical manner undermines the public's support for promising technologies and approaches. That government and public anxieties can quickly block research has been demonstrated repeatedly.

Take the nearly eight years of restrictions on human embryonic stem-cell research in the United States, instituted by President George W. Bush in 2001. Or the year-long voluntary moratorium called in 2012 on 'gain-of-function' experiments involving the highly pathogenic avian H5N1 influenza virus<sup>8</sup>. Here, researchers used genetic engineering to enhance the transmissibility of such flu viruses in mammals in the course of investigating changes that might increase their transmissibility between humans. The research is aimed at predicting which strains we shall need vaccines against in the near future.

Biology must move forward on safety and security. Let's not reinvent the wheel, but learn from those doing safety better. ■

**Tim Trevan** is a consultant on biosafety and biosecurity based in Damascus, Maryland, USA. He was formerly a diplomat with the UK government and the United Nations, dealing with biological weapons disarmament.

e-mail: [timtrevan@biosafety.co](mailto:timtrevan@biosafety.co)

1. CDC Report on the Potential Exposure to Anthrax (CDC, 2014).
2. Weick, K. E. & Sutcliffe, K. M. *Managing the Unexpected: Resilient Performance in an Age of Uncertainty* (Wiley, 2001).
3. Holland, D. & Shemwell, S. M. *Implementing a Culture of Safety: A Roadmap for Performance-based Compliance* (Xlibris, 2014).
4. Alcoa Annual Report 1998 (Alcoa, 1998); available at <http://go.nature.com/6odl3d>
5. 'The Peer Principle' *BusinessWeek* (May 2010).
6. Chassin, M. R. & Loeb, J. M. *Health Aff.* **30**, 559–568 (2011).
7. Oye, K. A. *et al. Science* **345**, 626–628 (2014).
8. Fouchier, R. A. M. *et al. Nature* **493**, 609 (2013).





Human embryos are prime targets for genome editing.

# A path through the thicket

As various advisory bodies, scientific organizations and funding agencies deliberate on genome editing in humans, **Debra J. H. Mathews, Robin Lovell-Badge** and colleagues lay out some key points for consideration.

**T**he ease of use, accuracy and efficiency of the genome-editing tool CRISPR/Cas9 has led to its broad adoption in research, as well as to preliminary applications in agriculture and in gene therapies involving non-reproductive (somatic) cells. It is also possible in some jurisdictions to deploy CRISPR/Cas9, and related techniques<sup>1</sup>, in human germline cells (sperm and eggs) as well as in early embryos<sup>2</sup>.

In September, a network of more than 30 scientists, ethicists, policymakers, journal editors and funders called the Hinxton Group gathered in Manchester, UK, to address the ethical and policy issues surrounding the editing of human genomes in the early stages of development and in germline cells (see [go.nature.com/xikxv2](http://go.nature.com/xikxv2)). Similar meetings have been and are being held elsewhere in the world, and several position statements have been published (see, for instance, [go.nature.com/enfxjz](http://go.nature.com/enfxjz) and [go.nature.com/fes1wc](http://go.nature.com/fes1wc)). Indeed, the US National Academies of Sciences, Engineering and Medicine is hosting what could be the largest such gathering next month, in concert with the Chinese Academy of Sciences and the Royal

## POINTS FOR DEBATE

Some key questions are emerging as advisory bodies across the world discuss the ethical and policy implications of genome editing in humans. Here are four:

- Should genome editing be allowed in basic research involving human sperm, eggs and embryos?
- Should only embryos left over from *in vitro* fertilization be used in genome-editing research or may embryos be specifically created for research?
- What safety and efficacy thresholds need to be met before the use of genome editing in human reproductive applications could be considered?
- If such thresholds are met, what uses for genome editing in human reproductive applications might be permissible?

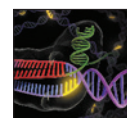
Society in London (see [go.nature.com/frauil](http://go.nature.com/frauil)).

Here, we lay out some key points emerging from the Hinxton meeting that are likely to gain more focus from the international community in the coming months. The views presented here are those of the group's steering committee and do not necessarily represent the consensus view of the group.

## RECOMMENDED ACTIONS

**Establish a model regulatory framework that could be adopted internationally.**

Various groups, including ours, agree that numerous technical and safety issues need to be addressed before genome-editing technologies could feasibly be used in reproductive clinical applications. Many also share our strong conviction that basic research involving genome editing should not be halted or hampered. Such studies are likely to have tremendous value, including in human-reproduction applications that do not ▶



**CRISPR GENE EDITING**  
A Nature collection  
[nature.com/crispr](http://nature.com/crispr)

► involve genome editing, and potentially in the development of treatments using somatic cells (see ‘Basic benefits’).

Even if much of the international scientific community and major funders of biomedical research agree on this point, however, the ease of use and accessibility of the technology make it ripe for exploitation by rogue or charlatan organizations — especially in jurisdictions where fertility clinics, which must be involved, are loosely regulated. After all, for the

**“We could not agree on the merits and downsides of listing possible clinical uses.”**

past decade, thousands of medical tourists have collectively paid many millions of dollars to receive unproven and unregulated stem-cell interventions internationally<sup>3,4</sup>.

To address this issue, it would be helpful to have a model regulatory framework that is specific enough to be meaningful, but general enough that it could quickly be adopted by country-specific regulatory bodies. There are precedents for such guidance; a document produced by the International Society for Stem Cell Research informs and influences oversight of stem-cell research internationally. But guidelines cannot in themselves prevent clinics from carrying out unproven treatments — that is the remit of local authorities. Much as the Human Fertilisation and Embryology Act, and the regulatory body called the Human Fertilisation and Embryology Authority (HFEA), have set limits around the use of embryos and gametes in research and the clinic in the United Kingdom, clear boundaries need to be established to indicate when research and clinical practice are permitted, with penalties given to those who step outside them. The public must also be made aware of what is legitimate and what is not, which will in turn require robust communication from authorities as well as from scientists and clinicians.

### Develop a road map for basic research.

Although much of the focus of public discussions has been on potential clinical applications, the immediate and perhaps most exciting uses of human genome editing are in basic research. Advisory and regulatory bodies should set priorities for genome-editing research involving human cells, including the germ line. This should involve canvassing a diverse group of scientists from across the world who have expertise on: genome-editing technologies; genomics and human genetic variation; mutation types, frequency and effects on physical and other traits; gene expression and regulation; epigenetics; human embryology and reproductive biology; and clinical genetics.

Two issues will be particularly crucial to address — both in the context of basic

research and especially in relation to the development of any human reproductive applications.

**Off-target events.** Computer-based and *in vitro* methods will be required to evaluate the likelihood of introducing changes outside the target site and the possible consequences. They will also be needed to distinguish such mutations from those that result naturally from, for example, imperfect DNA repair after cell division or environmental insults (chemicals, ionizing radiation and so on).

When genome editing is applied to cells that grow well in culture, such as spermatogonial stem cells (which give rise to sperm), a single edited cell can generate many millions of others. By sequencing the genomes of a subset of these cells, researchers can get a good indication of off-target events. Such assays are less reliable when only one or a few cells are available, as with cells biopsied from early embryos. In this case, one option would be to establish pluripotent cell lines from these cells. A key challenge will be to differentiate off-target events from the substantial natural sequence variation that exists between individuals and between cell lines.

Recent experiments<sup>5</sup> in mice and in human cell lines suggest that the rate of off-target events is insignificant compared with the number of spontaneous mutations that occur in each generation<sup>6</sup>. Yet the number of mutations may be less important than where they occur. Unlike spontaneous mutations, which are essentially random, off-target events are likely to be influenced by the RNA molecule that is used to guide the cutting component of the technology (the nuclease) to the right place in the genome. Also, most genome-editing research so far has been conducted in genetically homogeneous populations, such as inbred strains of mice. Little is known about what effect variation in genetic background might have on the efficiency and accuracy of genome editing, or on the ability of researchers to differentiate off-target events from background variation.

**Mosaicism.** When genome-editing tools are applied to multicellular embryos, only some of the cells may be altered, resulting in a genetic mosaic of edited and unedited cells. Even when applied to a single-cell embryo, the nuclease may not cut both copies of the target gene, or the cell may start to divide before the changes have been completed. It will be important to know what level of mosaicism is likely for different applications of genome editing, how this can be measured, and what effects mosaicism might have.

For some research purposes, such as tracking cell fate, mosaicism may not matter. (As long as some of the cells in the original sample carry the marker gene — say, the gene encoding green fluorescent protein — researchers will be able to identify what those cells develop into.) For others, mosaicism could be more problematic. For instance, if the aim is to determine the role of a gene whose product is secreted and the total amount of secreted product is crucial to gene function, then having a proportion of unedited cells will greatly affect the conclusions.

**Engage people from all sectors of society in a debate about genome editing, including the use of human embryos in this research.** Although human embryos and sperm and egg cells are used in a broad range of research internationally, including in the context of *in vitro* fertilization (IVF) and embryonic stem-cell research, some have suggested that their uses in genome-editing research should be considered independently. Importantly, embryos left over from IVF are unlikely to be a good model if mosaicism is to be avoided, because they comprise more than one cell; they are usually not made available for research until they are at or beyond the eight-cell stage of development. Also, founder effects, whereby most of a fetus derives from one or a few early cells, could complicate predictions about the functional consequences of mosaicism. Cell competition could have similar effects. Early analyses may indicate,

### BASIC BENEFITS

The use of genome editing in human sperm, eggs and embryos could yield valuable insights in several areas of basic research.

#### Research

How cell types are specified in the early human embryo, and the nature and importance of the genes involved.

Understanding the biology and genetics of stem-cell lines representing the cell lineages thought to exist in the early human embryo — including non-embryonic cells, such as those that go on to make the placenta.

The role of specific genes in human germ-cell development, including the differentiation of sperm and eggs.

Genome-editing techniques.

#### Possible applications

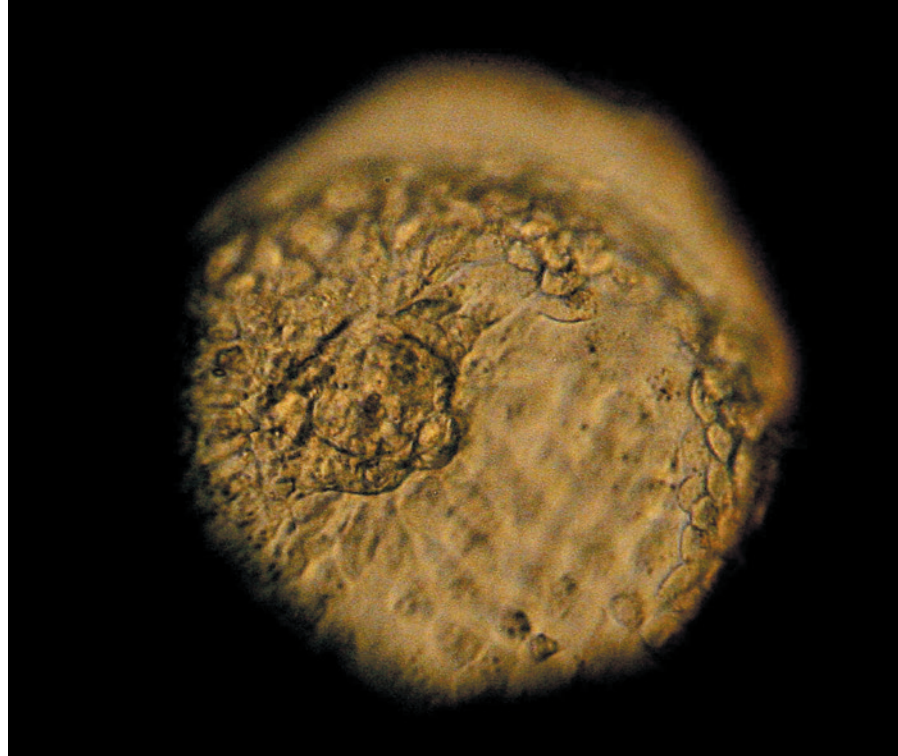
Improved techniques for culturing embryos following *in vitro* fertilization, better implantation rates and fewer miscarriages.

Improved ability to establish stem-cell lines for research, prevent miscarriages and screen drugs for efficacy. Reduction in the need for embryos in research.

Fertility enhancement and the development of new contraceptives.

Improved efficiency and versatility of genome editing in early embryos and in germline cells. Reduction in the numbers of embryos required in experiments.





Embryos donated to the La Jolla IVF clinic are made available for stem-cell research in the United States.

for instance, that even if 80% of the cells in an embryo were manipulated, the resulting fetus could turn out to contain a much smaller proportion of edited cells if such cells differ even very slightly in their rate of division or survival compared with unedited ones<sup>7</sup>.

Minimizing and so reducing the effects of mosaicism could require introducing the genome-editing components — the nuclease and the guide RNA — just after fertilization, or even at the point when fertilization happens<sup>8</sup>. If this is the case, work that involves editing the human germ line may end up being restricted to those jurisdictions that allow the creation of embryos specifically for research. Under current laws, this would limit studies to eight countries: Belgium, China, Israel, Japan, Singapore, South Korea, the United Kingdom and the United States (R. Isasi, personal communication). And in the United States, such work would be possible only with the use of non-federal funds.

If it turns out that major benefits arise from basic research in human-genome editing — including applications that do not in themselves involve genome editing, such as ways to improve fertility or reduce the incidence of miscarriage — then national laws could jeopardize people's access to such benefits.

**Design tools and methods to enable inclusive and meaningful deliberation.** Groups such as the UK Nuffield Council on Bioethics and the national academies in the United States and elsewhere are well placed to take the lead on efforts to ensure that debates over the use of human embryos in basic research are geographically and demographically inclusive, and that such debates inform policy decisions. This is also crucial in relation to the even thornier question of what clinical applications might be appropriate, given

sufficient assurance of safety and efficacy.

Even for members of the Hinxton Group — well-informed people who concur that basic research involving genome editing, including that in human sperm, eggs and embryos, has tremendous value — discussions about possible reproductive applications were charged. In fact, we found it impossible to agree on potentially defensible uses for editing the human germ line without having the context and facts of a particular case. The group considered a spectrum of interventions, from the correction of life-threatening mutations (those that cause Tay-Sachs disease, cystic fibrosis and Huntington's disease, for example), to the introduction of preventive changes (including disruption of the CCR5 gene to confer resistance against HIV infection), to non-medical enhancements (such as increasing muscle mass). Ultimately, we could not even agree on the merits and downsides of listing possible clinical uses — some were concerned that making such a list could be taken as tacit approval.

For decades, people have been arguing about the pros and cons of human germline modification<sup>9</sup>, how to distinguish medical treatments from enhancement, what rights parents have over the lives of their children and so on. Yet good models for how to enable a diversity of perspectives to shape morally contested areas of emerging science and technology are hard to find.

When the US National Institutes of Health (NIH) published draft guidelines for embryonic stem-cell research in 2009, nearly 50,000 people sent in comments. Many of those who weighed in were non-scientists, including about 16,000 who opposed the research on moral grounds. On seeing that their comments had failed to shape the final guidelines, opponents of stem-cell research

went to the courts, and four years of litigation and uncertainty for the field followed.

More recently, the HFEA conducted a programme of public engagement that included workshops, focus groups and online surveys to gauge opinion on a range of issues relevant to the use of mitochondrial replacement therapy (MRT) in the clinic. (In MRT, which was approved by the UK government in February, faulty mitochondrial DNA in an egg or embryo is discarded and DNA from a woman without mitochondrial disease is used.) This programme largely achieved its goals and is one of the best examples we have of such an effort (see [go.nature.com/64cioj](http://go.nature.com/64cioj)). To some, both of these examples of public engagement constitute successes, but we have no a priori measures to indicate what success means.

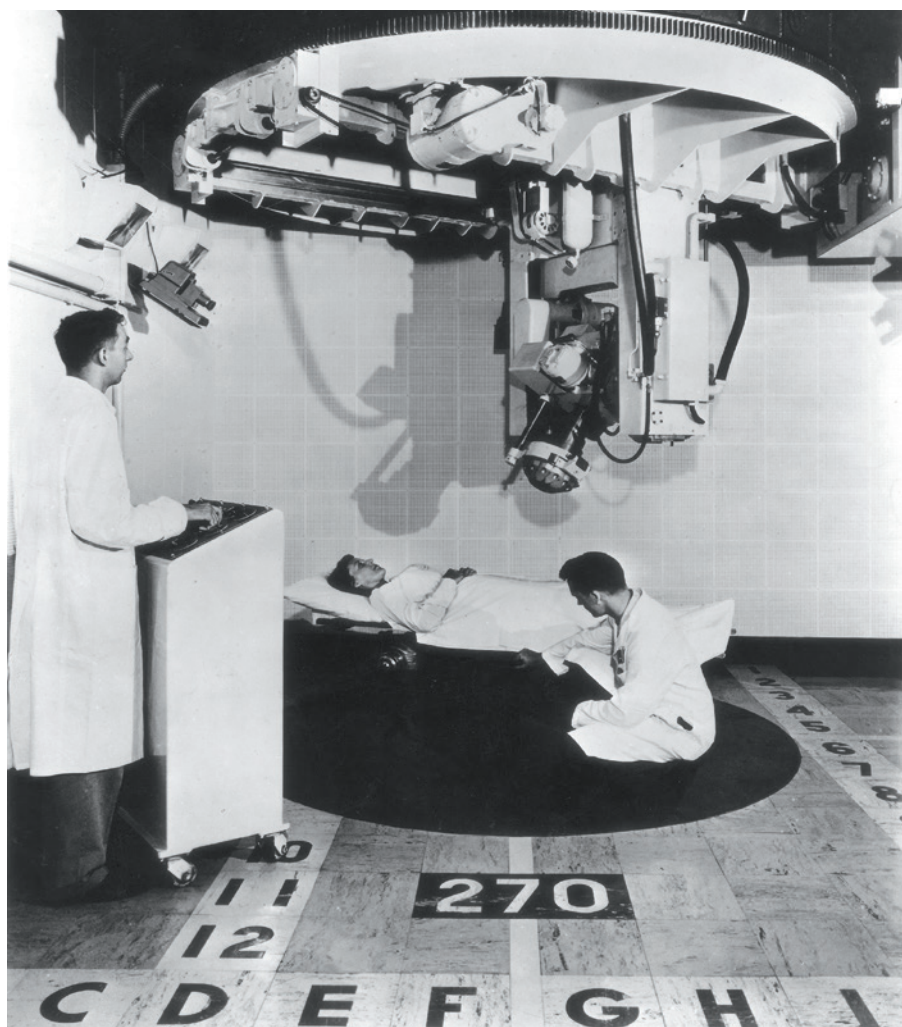
Systematic study of various models of engagement is needed to identify, for instance, the best methods for ascertaining not only partisan positions, but also broad societal attitudes. So too are investigations that probe how public deliberations can shape policy.

The advances in genome-editing technologies mean that long-standing ethical questions can no longer be dodged on the basis of obvious and agreed safety concerns. And although continued debate is crucial, it is time to collectively make decisions about the kind of world we want to live in and to develop policies to reflect that vision. ■

**Debra J. H. Mathews** is assistant director for science programmes at the Johns Hopkins Berman Institute of Bioethics, Baltimore, Maryland, USA, and associate professor in the Department of Pediatrics at the Johns Hopkins University School of Medicine. **Robin Lovell-Badge** is group leader at the Laboratory of Stem Cell Biology and Developmental Genetics, the Francis Crick Institute, London, UK. The rest of the Hinxton Group Steering Committee are **Sarah Chan, Peter J. Donovan, Thomas Douglas, Christopher Gyngell, John Harris and Alan Regenberg**. e-mail: [dmathews@jhu.edu](mailto:dmathews@jhu.edu)

1. Zetsche, B. *et al.* *Cell* **163**, 759–771 (2015).
2. Liang, P. *et al.* *Protein Cell* **6**, 363–372 (2015).
3. Regenberg, A. C., Hutchinson, L. A., Schanker, B. & Mathews, D. J. H. *Stem Cells* **27**, 2312–2319 (2009).
4. Einsiedel, E. F. & Adamson, H. *Dev. World Bioeth.* **12**, 35–44 (2012).
5. Tan, E. P., Li, Y., Velasco-Herrera Mdel, C., Yusa, K. & Bradley, A. *Genesis* **53**, 225–236 (2015).
6. Shendure, J. & Akey, J. M. *Science* **349**, 1478–1483 (2015).
7. Sancho, M. & Rodríguez, T. A. *Cell Cycle* **13**, 9–10 (2014).
8. Suzuki, T., Asami, M. & Perry, A. C. F. *Sci. Rep.* **4**, 7621 (2014).
9. US President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. *Splicing Life: The Social and Ethical Issues of Genetic Engineering with Human Beings* (US Government Printing Office, 1982).

The authors declare competing financial interests. For details, and for full author affiliations, see [go.nature.com/bhrwgt](http://go.nature.com/bhrwgt).



A woman has cancer radiotherapy at Oak Ridge Institute of Nuclear Studies, Tennessee, in the 1950s.

## CANCER THERAPY

# Up close and personalized

Gerard Evan reviews an inspiring, at times frustrating, chronicle of the war on cancer by one of its generals.

The 1971 US National Cancer Act, which effectively declared a 'war on cancer', was the product of a confused era. During the cold war, extreme pessimism and extreme optimism were constant, if unlikely, bedfellows. As long as we didn't blow ourselves up, anything was possible — even curing cancer. In the years since, Vincent DeVita has been both a key participant in and a witness to this war, as director of the US National Cancer Institute (NCI), physician-in-chief at Memorial Sloan Kettering Cancer

Center in New York City and director of the Yale Cancer Center in New Haven, Connecticut. In *The Death Of Cancer*, he and his daughter, the accomplished science journalist Elizabeth DeVita-Raeburn, write of this period with passion and insight. It is a deeply personal account, often inspiring, sometimes confusing and occasionally unifying.

DeVita opens with the story of his friend Lee. Lee's protracted encounter with aggressive prostate cancer encapsulates the book's principal thesis: that the treatment of cancer

is frequently ill-informed, too conservative and mired in the timidity of local and national bureaucracies. DeVita makes an unassailable case that advances come from people with courage, vision and tenacity, rare virtues too often confused with risk-taking and stubbornness in today's blame-addicted society.

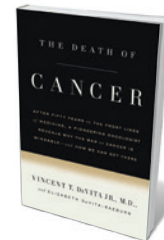
Lee's cancer follows a well-trodden course, initially responding to therapy, but over time, and under the selective pressure of drugs, evolving into a refractory, terminal disease. DeVita uses his abundant connections to see that Lee receives exceptional personalized care — radiotherapy in the best cancer hospitals and unprecedented access to experimental drugs. Through this heroic effort, Lee survives for 12 years. Yet with this narrative, DeVita also opens a can of worms. He shows that the standard of care can be improved. But a uniquely well-connected ex-head of the NCI personally stewarding someone's treatment is personalized medicine for the super-elite (and the cost be damned).

It is also irritating to see DeVita fall into the teleological trap of endowing cancer with baleful purpose. He describes cancer cells as "canny" and "smart" — they "learn to outwit" therapy and "figure out" ways of surviving. Poppycock. Cancer progression and the emergence of resistance are the results of goal-less evolutionary processes, not calculated intent. Imbuing cancers with malign purpose terrifies patients and is shoddy biology.

After Lee, we enter the barbaric world of post-First World War cancer treatment, when crude surgery and even cruder radiotherapy ruled. We then meet JD, a Polish immigrant with lymphoma and the first person to get chemotherapy for cancer. In 1942, he was admitted to what is now Yale–New Haven Hospital, where, having previously failed radiotherapy, he was pumped with nitrogen mustard (also known as mustine). His cancer shrank to nothing, but the chemotherapy killed him. From such an unprepossessing start, it took guts to bear the chemotherapy torch and, given the toxicity of the agents, even more guts to use them in combination.

DeVita tells an engaging tale of how combination chemotherapy was developed at the NCI, and of his pivotal role. The NCI,

a quasi-independent juggernaut of the National Institutes of Health, was tasked with overseeing prosecution of the National Cancer Act. We meet the remarkable Emil 'Tom' Frei, NCI chief of medicine, and Emil 'Jay' Freireich, who ran the children's leukaemia ward. They developed the first cure for childhood leukaemia,



**The Death of Cancer**  
VINCENT T. DEVITA  
AND ELIZABETH  
DEVITA-RAEBURN  
Sarah Crichton: 2015.



and coined the first of the snazzy acronyms: VAMP, a brew of vincristine, amethopterin, 6-mercaptopurine and prednisone. In 1972, DeVita, Frei and Freireich were among the recipients of the Albert Lasker Clinical Medical Research Award for their work.

They did not have an easy time of it. DeVita paints a depressing picture of internecine strife in oncology, fuelled by a preoccupation with status and status quo at the expense of innovation. Be warned: this is a very personal tale of political manoeuvring, with a large cast of heroes and anti-heroes. I recommend noting down the names as you go.

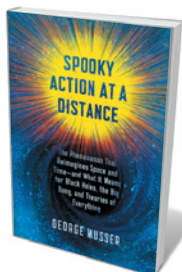
DeVita became director of the NCI in 1980. He had the thankless task of replacing a grace-and-favour establishment with something approaching a meritocracy — guaranteed to elicit some lifelong enmity. He also presided over the maturation of combination chemotherapy from an empirical mix of poisons into a modern, evidence-based process, undoubtedly saving millions of lives.

Yet here, as with the section on Lee (and for much the same reason), DeVita's narrative occasionally raised my hackles. The 1960s and 1970s were complex and difficult: a dearth of cancer drugs, mechanisms and vision was seasoned with vested interests and institutional insecurity. But DeVita describes a world where influential doctors used their power to demand favours of each other and their underlings — a system in which he participated. This is not a reassuring depiction of medical practice. Of course, it is impossible not to admire the passion, optimism and dedication that course through this book, and I applaud every success among patients for whom DeVita cared personally. But we must never forget that the best cancer therapies need to be made available to everyone, at a cost that patients and society can afford.

The war on cancer is often compared unfavourably to another great US science initiative, the Apollo lunar programme. This is unfair. Both required almost superhuman ingenuity and effort, but landing a man on the Moon was principally a problem of engineering and organization. By contrast, despite 45 years of insight into the molecular processes of cancers, we still do not really know how therapies exploit cancers' vulnerabilities, or even why such vulnerabilities exist. Every day, we use targeted drugs to block mutations that drive cancer cells but, remarkably, we still fail to understand why that kills them. It is ironic, then, that former US president John F. Kennedy's prophetic words about the Moon mission best describe our enduring struggle against cancer: "The greater our knowledge increases, the greater our ignorance unfolds." ■

**Gerard Evan** is Sir William Dunn Professor of Biochemistry at the University of Cambridge, UK.  
e-mail: gie20@cam.ac.uk

## Books in brief



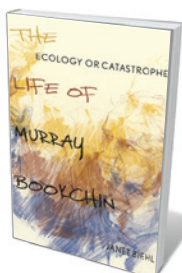
### Spooky Action at a Distance

**George Musser** SCIENTIFIC AMERICAN/FARRAR, STRAUS AND GIROUX (2015)  
Does space exist? The quantum phenomenon of non-locality, in which two particles can be correlated even when far apart, begs that question. In this polished study of the concept that Albert Einstein dubbed "spooky action at a distance", science writer George Musser tours the entangled research, history and philosophical speculation surrounding it. He examines the heated debates (such as the recent wrangle between the string and twistor communities), the theories of physicists such as laser pioneer Enrique Galvez and more, proving that this is one of the most engrossing disputes in science.



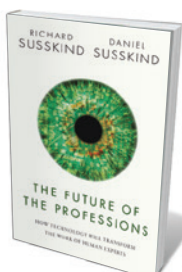
### Concrete Revolution: Large Dams, Cold War Geopolitics, and the US Bureau of Reclamation

**Christopher Sneddon** UNIVERSITY OF CHICAGO PRESS (2015)  
In this stellar history, geographer Christopher Sneddon traces the twentieth-century boom that saw 50,000 big dams built worldwide. The US Bureau of Reclamation presided, from the Great Depression megaproject Hoover Dam to the cold-war export of bureau engineers to more than 100 countries. Yet by 1969, assistant commissioner Gilbert Stamm saw that doing "marvellous things with materials" does not necessarily meet human needs. Societies and rivers, Sneddon shows, make for a complex confluence.



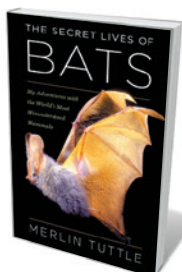
### Ecology or Catastrophe: The Life of Murray Bookchin

**Janet Biehl** OXFORD UNIVERSITY PRESS (2015)  
Just before the explosive advent of Rachel Carson's book *Silent Spring* in 1962, radical ecologist Murray Bookchin published *Our Synthetic Environment*, warning of impacts from pesticides, industrial farming and deforestation. He went on to pen *Crisis in Our Cities* (1965), which predicted global warming from fossil-fuel use, and to teach solar-power technology and urban farming. The prescient Bookchin emerges in Janet Biehl's politics-heavy biography as incisive, inventive and pragmatic — a refreshing contrast to today's environmental doom-mongers and techno-utopians alike.



### The Future of the Professions: How Technology Will Transform the Work of Human Experts

**Richard Susskind and Daniel Susskind** OXFORD UNIVERSITY PRESS (2015)  
Advances in digitization will soon obviate the need for doctors, teachers and lawyers. So write 'legal futurist' Richard Susskind and economist Daniel Susskind, arguing that the professions are unaffordable, antiquated and opaque. Analysing how the algorithmic juggernaut is forcing the decomposition of traditional careers, the authors propose six new professional models — such as "knowledge engineers" — that together form a non-alarmist vision of how "increasingly capable" machines could help to redistribute expertise.



### The Secret Lives of Bats: My Adventures with the World's Most Misunderstood Mammals

**Merlin Tuttle** HOUGHTON MIFFLIN HARCOURT (2015)  
Anyone who has ever thrilled to bats' aerial feats, pollinating prowess or outright charisma will delight in ecologist Merlin Tuttle's scientific memoir. Drawing on 55 years of research, Tuttle relates exhilarating moments galore, from his teenage sighting of a "mother lode" of 100,000 gray myotis deep in a Tennessee cave, to dangling from a hot-air balloon to spot free-tail bats and dodging spears in Kenya while hot on the trail of chiropteran carnivores. **Barbara Kiser**



Katharina Kepler depicted here being threatened with torture.

#### HISTORY OF SCIENCE

## Trial by gender

**Jennifer Rampling** applauds an account of how Johannes Kepler saved his mother from being burned as a witch.

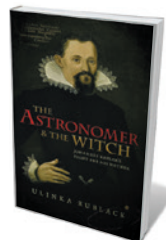
On the cover of *The Astronomer and the Witch* is a portrait of Johannes Kepler: confident, well dressed, half-smiling. This is the image that the imperial mathematician projected to his peers and patrons. But it is not the Kepler whom we meet in Ulinka Rublack's enthralling book — anxious, harassed by financial and family problems, and not even the main character. Rublack's protagonist is the great astronomer's mother, the unfairly accused "witch" of the title. In placing Katharina centre stage, Rublack tells a new story, one that is as much social history as it is scientific revolution.

As women in early modern Europe aged and their fertility declined, so did their status. In a small town such as Leonberg in the duchy of Württemberg, part of present-day Germany, even a respectable property owner like Katharina Kepler could not escape the stigma of age. Yet women did not all face this challenge equally. Rublack juxtaposes Katharina's hard life with the luxurious retirement of Sibylle, widowed duchess of Württemberg, on whom, in a way, her fate came to depend.

While Katharina struggled to work her land and raise children, often without their father, the rough ground behind Leonberg Castle was cleared to provide Sibylle with spectacular gardens. Investing in medicinal

plants added to the duchess's standing as a patroness of the poor and sick. Katharina, tired and short-tempered, did not enjoy the same indulgence. Her use of herbal remedies, common enough at the time, raised suspicion after a local woman blamed her illness on a "witches' brew" served by Katharina. After more allegations, and many delays, in 1619 Katharina was formally accused of witchcraft.

What distinguishes Katharina's case from thousands like it is the involvement of her famous son — the reason that the trial records have been preserved. Sifting through these,



**The Astronomer and the Witch: Johannes Kepler's Fight for his Mother**  
ULINKA RUBLACK  
Oxford University Press: 2015.

Rublack reconstructs an atmosphere of anxiety and suspicion as Württemberg slid into the Thirty Years' War. Accusations of witchcraft threw whole families under suspicion, and the taint of religious unorthodoxy could damage careers, as Johannes Kepler discovered when his Calvinist leanings blocked him from a post at the University of Tübingen. As Rublack notes,

to win patronage required "not only a powerful intellect and vision, but also books, ink, piety, and perfect manners". Having mastered these resources in his precarious career as mathematician to three emperors, Kepler now deployed them on behalf of his family. After all, he had achieved fame promoting a controversial position, Copernicanism, for which standards of proof were considerably higher than those required to burn a witch.

Rublack shows how Kepler bent his experience towards deconstructing faulty arguments and marshalling evidence for Katharina's defence. The legal system provided checks and balances, for example by requiring multiple witnesses, but in practice short cuts were taken. The case against Katharina was assembled by administrators sympathetic to her accusers, and witnesses gave conflicting evidence or described events from their childhoods. To this might be added, at any time, a 'confession' under torture. Convicted witches were usually burned alive.

In this fraught environment, Kepler repeatedly revised his *Harmony of the World* (1619): a five-part magnum opus that included his third law of planetary motion and his views on topics as diverse as musical theory and astrology. In a chapter on psychology, Kepler betrayed his own ambiguous feelings on his mother's plight, asking whether she had brought her misfortune on herself. Yet, by the time he prepared her final defence two years later, the astronomer had devised a new narrative: one that granted old women such as Katharina a role as knowledge-makers.

Rublack argues that Kepler justified his mother's medical practice by drawing a parallel with privileged women such as Sibylle. He claimed that women's experience and observation, gained (often painfully) over long periods of time, "constituted a basis for reputable and probable, if not certain knowledge". Kepler defended his mother by using Sibylle as an unimpeachable role model of a pious woman dispensing medical care.

Kepler wins the day, for although Katharina spoke in her own defence, it is her son's arguments that are preserved verbatim in the trial documents, commenting on and judging women's behaviour. Chained in her cell at the centre of the controversy, Katharina's own voice is harder to hear. Rublack calls out Kepler's past biographers for dismissing his mother as quarrelsome, difficult, "witch-like". If I have one criticism of the book, it is that its title plays to that stereotype, rather than to the nuanced characterization that the author has drawn. Rublack's vigorous, early modern anti-heroine was, surely, entitled to her anger. ■

**Jennifer Rampling** is assistant professor of history at Princeton University in New Jersey, where she teaches the history of early modern science.  
e-mail: [rampling@princeton.edu](mailto:rampling@princeton.edu)



# Correspondence

## Lessons from Madrid for next climate talks

As the 2015 Paris climate summit approaches, it is worth noting that this month marks 20 years since the Madrid meeting at which the Intergovernmental Panel on Climate Change (IPCC) declared that “the balance of evidence suggests a discernible human influence on global climate”.

Support for the 1995 IPCC statement came from the physical understanding of heat-trapping properties of greenhouse gases, observations of warming and comparisons of modelled and observed climate-change patterns (climate ‘fingerprinting’).

Criticism at the time stemmed from the paucity of fingerprint studies, the inadequate treatment of uncertainties, the focus on surface temperature, and poor quantification of natural climate noise and human-caused warming.

One lesson from Madrid is the importance of responding to justifiable criticism. Climate forensics have since identified human-caused temperature fingerprints from the stratosphere to the ocean depths, and in many variables other than temperature. It is routine to assess uncertainties in observed climate data and model simulations. The human-caused warming signal and the noise of natural climate variability are now better quantified. This signal has dominated since the mid-twentieth century.

We have also learned that global scientific understanding can emerge in less than 20 years from the noise of unreason and disinformation, and that one sentence can change the world.  
**Benjamin D. Santer** *Lawrence Livermore National Laboratory, California, USA.*  
santer1@llnl.gov

## Account for soil as natural capital

Economics has an important connection with the health of soils (see, for example, P. Panagos

*et al. Nature* 526, 195; 2015).

Integrating information on soil resources with other measures of natural capital and economic activity remains one of the least developed areas of the United Nations System of Environmental Economic Accounting (SEEA).

The SEEA is a broad-scale monitoring tool that is gaining global momentum. It integrates environmental data with economic measures such as national income, stock markets and gross domestic product.

The potential of the SEEA and natural-capital accounting to support regional, national and global monitoring efforts is being rapidly recognized in forums such as the UN Sustainable Development Goals, the Aichi Biodiversity Targets and the development of a Natural Capital Protocol for businesses. Accounting for soil resources makes a valuable contribution to this bigger picture.

**Carl Obst** *University of Melbourne, Australia.*  
cobst@unimelb.edu.au

## Drug pollution: Europe responds

As directors of the Association of the European Self-Medication Industry (AESGP), the European Federation of Pharmaceutical Industries and Associations (EFPIA), and the European Generic and Biosimilar Medicines Association (EGA), respectively, we are committed to providing safe, efficient and high-quality medicines without releasing harmful components into the environment (see *Nature* 526, 164; 2015).

Effluents from drug manufacturing account for just 2% of the pharmaceuticals found in the environment in Europe (see [go.nature.com/ovgyaa](http://go.nature.com/ovgyaa)) because they are managed effectively (D. J. Caldwell *et al. Environ. Toxicol. Chem.* <http://doi.org/8xf>; 2015). The industry has started to control drug pollution from other sources

in response to legislation that governs all aspects of pharmaceutical operations. Monitoring continues even after medicines are on the market.

This is complemented by such initiatives as our Eco-Pharmaco-Stewardship framework, a holistic environmental risk-management programme. We also run a joint medicines-disposal campaign on social media ([www.medsdisposal.eu](http://www.medsdisposal.eu)). And the iPIE project of the Innovative Medicines Initiative uses targeted assessment to identify the environmental risks of active drug contaminants.

**Hubertus Cranz** *AESGP, Brussels, Belgium.*

**Richard Bergström** *EFPIA, Brussels, Belgium.*

**Adrian van den Hoven** *EGA, Brussels, Belgium.*  
[richard.bergstrom@efpia.eu](mailto:richard.bergstrom@efpia.eu)

## Drug pollution: industry responds

The multinational drug company AstraZeneca has a long-standing commitment to proactively manage the environmental risks associated with drug manufacture and products in patient use (see *Nature* 526, 164; 2015). In the absence of legislation to control discharges from production sites, we have voluntarily devised concentration limits to ensure the safety of aquatic environments.

These environmental reference concentrations and maximum tolerable concentrations are derived using approaches similar to environmental quality standards developed under the Water Framework Directive (R. J. Murray-Smith *et al. Integr. Envir. Assess. Manag.* 8, 320–330; 2012). We have established them for all 42 of the company's active pharmaceutical ingredients.

Our global manufacturing sites and our key outsourced manufacturing partners meet these safe-discharge criteria. This should help to raise environmental standards across the industry. We completed 72 assessments of environmental

reference concentrations with our supply chain in 2014.

We continuously collate published environmental data on all of our active pharmaceutical ingredients to keep our risk assessments and safe-discharge standards scientifically up to date (G. Holm *et al. Drug Safety* 36, 533–546; 2013).

**Jason Snape, Wesley White** *AstraZeneca, Alderley Park, UK.*  
[jason.snape@astrazeneca.com](mailto:jason.snape@astrazeneca.com)  
*Competing financial interests declared; see [go.nature.com/uckccxf](http://go.nature.com/uckccxf).*

## Irked by naivety about policymaking

We find William Sutherland and Mark Burgman's advice on the complex social processes between evidence and policy decisions to be naive at best, and antidemocratic at worst (see *Nature* 526, 317–318; 2015).

Policymakers are influenced by a much greater range of factors than are considered by the authors — including public opinion, inheritance of policies and institutional rules, finance, unpredictable events, and trust in actors (see also W. Pearce *et al. Evid. Policy* 10, 161–165; 2014).

In our view, the authors perpetuate negative stereotypes of policymakers and academics, when in fact many examples of productive collaborations and hybrid roles exist. Their edicts seem to undermine colleagues who mobilize knowledge for policy, and to reduce the intricate relationship between evidence and policy to a linear, technocratic process. As they themselves attest, giving advice to policymakers or academics that is not evidence-based could hamper the formation of useful collaborations.

**Kathryn Oliver** *University of Oxford, UK.*

**Adam Wellstead** *Michigan Technological University, Houghton, USA.*

**Paul Cairney** *University of Stirling, UK.*  
[kathryn.oliver@spi.ox.ac.uk](mailto:kathryn.oliver@spi.ox.ac.uk)

## MOLECULAR BIOLOGY

## DNA repair without flipping out

Time-resolved molecular snapshots of the bacterial enzyme AlkD reveal an unprecedented mechanism for the recognition and removal of damaged bases in DNA, with implications for cell biology and cancer therapy. [SEE LETTER P.254](#)

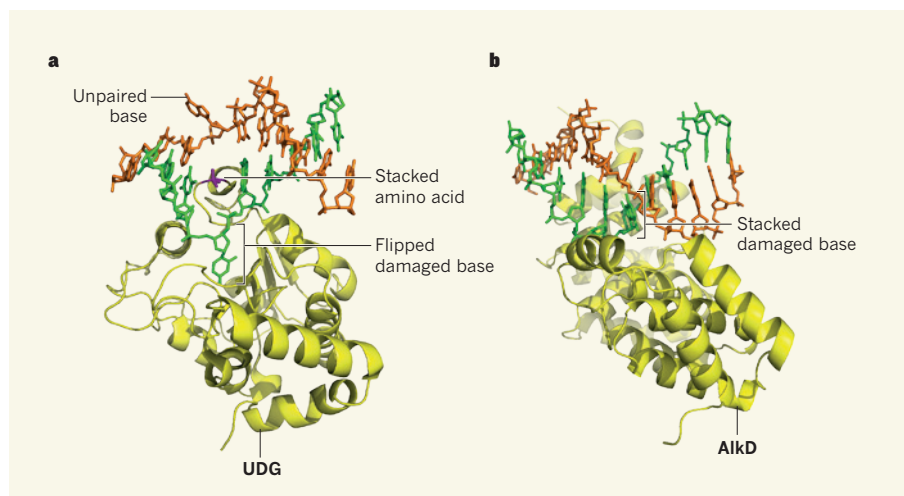
DAVID S. SHIN & JOHN A. TAINER

Many people imagine DNA as a highly stable blueprint that allows offspring to inherit the traits of their parents. However, DNA is a dynamic molecule that is not only read, replicated and modified by various cellular factors, but is also constantly being damaged. DNA repair mechanisms are crucial for maintaining sufficient sequence fidelity for cell survival and disease prevention. In October 2015, Tomas Lindahl, Paul Modrich and Aziz Sancar shared the Nobel Prize in Chemistry for their descriptions of DNA repair processes, which all involve enzymatic opening of the DNA double helix and unpairing of one or more DNA bases. In this issue, Mullins *et al.*<sup>1</sup> (page 254) describe a new class of DNA base-repair enzyme that unexpectedly leaves the double helix largely intact during the first step of damage recognition and removal.

Every human cell contains around 3 billion nucleotides — each consisting of one of four bases, along with a sugar and a phosphate molecule — that join together to form DNA. This structure was assumed to be unusually stable until Lindahl challenged this idea when he found that DNA was subject to decay (see Lindahl's review<sup>2</sup>). DNA can be damaged by environmental effects, such as ionizing radiation, ultraviolet light and chemicals, and by internal sources, such as oxygen radicals, alkylating agents or even hydrolysis by water. Lindahl also discovered DNA glycosylase enzymes (which remove damaged DNA bases, referred to as lesions) and was the first to describe protein interactions with DNA for base repair<sup>2</sup>.

Forms of base damage include deamination (loss of an amine group), oxidation, the attachment of alkyl or larger groups, and spontaneous detachment from the DNA backbone, which leaves behind toxic abasic sites termed apurinic or apyrimidinic sites<sup>2</sup>. All such lesions are predominantly repaired by the base-excision repair pathway<sup>3,4</sup>, in which a specific glycosylase binds to and bends DNA, opening the minor groove of the helix to remove the lesion, followed by replacement of the nucleotide by DNA-repair proteins.

Given the variety of lesions and the fact that bases are largely buried in the double-helix



**Figure 1 | Two mechanisms for base excision.** Damaged DNA bases represent potential loss of fidelity and toxicity to a cell. **a**, A well-described mechanism for the excision of damaged bases is that enacted by glycosylase enzymes such as uracil DNA glycosylase (UDG)<sup>2,3</sup>. This process involves substantial conformational change in the DNA, which elicits unpairing of DNA bases such that the damaged base is 'flipped out' of the DNA double helix (DNA strands are shown in green and orange). The glycosylase enzyme then inserts one of its own amino acids into the helix to stabilize the remaining base stack and prevent the flipped base from returning before the enzyme's catalytic process has removed it. **b**, Mullins *et al.*<sup>1</sup> show that the glycosylase AlkD achieves excision of damaged bases without base flipping and without gross rearrangements to the DNA. How the excised base, which remains buried in the DNA helix, exits the DNA after cleavage remains unclear.

structure, how are damaged bases specifically recognized and removed? Crystal structures have revealed that the whole nucleotide is 'flipped out' of the DNA helix and sequestered in the active site of the enzyme for repair<sup>5,6</sup> (Fig. 1a). During this process, an amino acid from the enzyme moves into the base stack to stabilize the base-flipped DNA intermediate<sup>3,6</sup>. The geometric and chemical complementarity between the enzyme's active site and the damaged base provides specificity for the reaction, and the positioning leaves the glycosidic bond that links the damaged base to its sugar group accessible to cleavage by the enzyme.

Mullins *et al.* present a mechanism for base recognition and removal that forgoes base flipping and amino-acid intercalation. The authors studied the alkylpurine DNA glycosylase (AlkD), which binds to adenine and guanine bases that have been modified by alkylation in the bacterium *Bacillus cereus*. These bases are prone to spontaneous detachment, leaving abasic sites, and their presence

may inhibit DNA replication<sup>7</sup>. The interactions of enzymes with alkylated bases is of great interest — control of DNA alkylation is central to the epigenetic marks that enable cells to differentiate into distinct tissues and to have distinct protein-expression profiles. And unrepaired alkylation damage can lead to mismatched bases and even cell death without successful mismatch repair, which was the focus of Paul Modrich's Nobel-prizewinning work. Furthermore, some cancer therapies are designed to induce alkylation damage with the aim of overwhelming aberrant repair mechanisms in tumour cells<sup>8</sup>. Indeed, future therapies may target specific repair responses as a 'weak spot' for many cancer cells<sup>9</sup>.

Previous X-ray crystal structures have shown AlkD to have a different architecture from other glycosylases, being composed of repetitive structures (HEAT repeats) that cradle DNA in a concave channel<sup>10</sup>. Mullins *et al.* now provide structures of the enzyme 'flash-cooled' at different times as it excises alkylated



bases. These structures show that, although the DNA does bend and has slight minor-groove widening, the distortion is mild in comparison with DNA bound to other glycosylases, and the modified nucleotide is not flipped out and sequestered in an active-site pocket (Fig. 1b).

Instead, the authors show that much of the AlkD–DNA interaction is provided by multiple hydrogen bonds between the charged protein and the DNA phosphate backbone that flanks the lesion. Furthermore, two tryptophan amino-acid residues in the enzyme contribute to both binding to the lesion in the minor groove and enzyme catalysis, in what are called CH– $\pi$  interactions. Such interactions are known from other protein–ligand recognition processes, but this may be the first example of CH– $\pi$  interactions that directly function in DNA-repair catalysis. Thus, Mullins and colleagues' findings reveal new chemistry as well as an unprecedented damage-recognition mechanism.

For clinical applications directed at the DNA-damage response, we now face the possibility that humans and pathogens have enzymes that sense and remove alkylated DNA lesions without binding pockets, and thus require a different mode of targeting from other known glycosylase reactions. In fact, some bacteria produce an alkylating agent, yatakemycin, as an antibiotic, and protect themselves from their own alkylating toxin using a glycosylase similar to AlkD<sup>11</sup>. Mullins *et al.* found that the large yatakemycin molecule fits into a solvent-filled gap between AlkD and DNA and could also be excised from this position by AlkD. This finding highlights the remarkable properties of repair without base flipping — the ability to recognize and remove small base modifications in DNA and to have activity against large base adducts on the outside of DNA. Furthermore, the mechanism requires little conformational change in either the enzyme or the DNA; because conformational changes use energy and are often the rate-limiting step of enzymatic reactions, AlkD may act more rapidly than other glycosylases.

Mullins and colleagues' report expands our knowledge of recognition and excision between proteins and alkylated bases. However, we are left with questions about how well the AlkD mechanism works and which cells might use similar protein interactions with alkylated DNA. The damage-recognition and -removal mechanism of AlkD seems less specific than the flipping mechanism of other glycosylases, although the AlkD–DNA interface provides some specificity by restricting AlkD to removing inherently labile, positively charged lesions. In humans, base flipping by glycosylases helps to coordinate 'hand-off' to an endonuclease enzyme that cuts the DNA backbone at abasic sites<sup>12</sup>; it will be interesting to see whether AlkD can perform a similar hand-off. AlkD may also have roles in the removal of bulky lesions, which was the focus

of Aziz Sancar's Nobel-prizewinning work. Thus, despite the earlier breakthrough research on DNA repair highlighted by this year's Nobel prize, Mullins and colleagues' findings on AlkD show that we still have much to learn. ■

**David S. Shin and John A. Tainer** are in the Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. J.A.T. is also in the Department of Molecular and Cellular Oncology, University of Texas MD Anderson Cancer Center, Houston.  
e-mails: davidsshin@lbl.gov;  
jtainer@mdanderson.org

1. Mullins, E. A. *et al.* *Nature* **527**, 254–258 (2015).
2. Lindahl, T. *Nature* **362**, 709–715 (1993).
3. Hitomi, K., Iwai, S. & Tainer, J. A. *DNA Repair* **6**, 410–428 (2007).
4. Lindahl, T. & Wood, R. D. *Science* **286**, 1897–1905 (1999).
5. Roberts, R. J. & Cheng, X. *Annu. Rev. Biochem.* **67**, 181–198 (1998).
6. Slupphaug, G. *et al.* *Nature* **384**, 87–92 (1996).
7. Larson, K., Sahm, J., Shenkar, R. & Strauss, B. *Mutat. Res.* **150**, 77–84 (1985).
8. Tubbs, J. L., Pegg, A. E. & Tainer, J. A. *DNA Repair* **6**, 1100–1115 (2007).
9. Farmer, H. *et al.* *Nature* **434**, 917–921 (2005).
10. Robinson, E. H., Gowda, A. S., Spratt, T. E., Gold, B. & Eichman, B. F. *Nature* **468**, 406–411 (2010).
11. Xu, H. *et al.* *Angew. Chem. Int. Edn* **51**, 10532–10536 (2012).
12. Mol, C. D., Izumi, T., Mitra, S. & Tainer, J. A. *Nature* **403**, 451–456 (2000).

This article was published online on 28 October 2015.

## ASTRONOMY

## A small star with an Earth-like planet

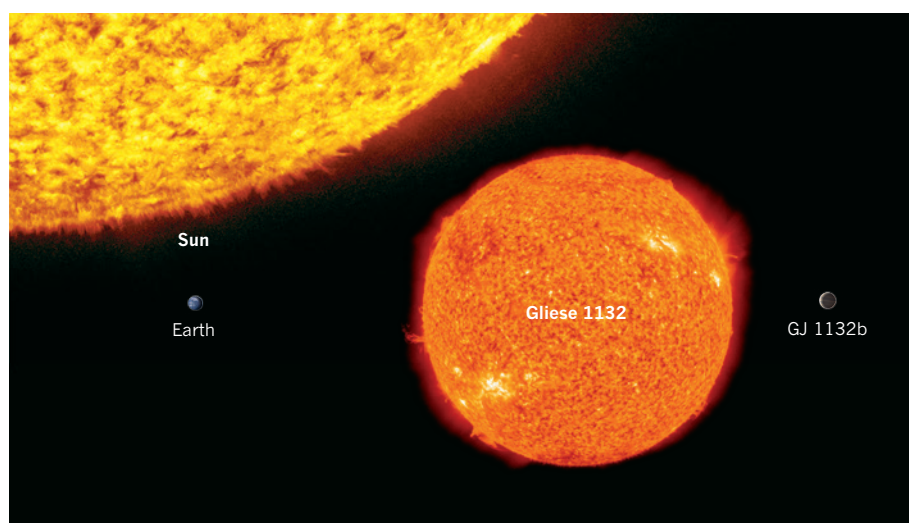
A rocky planet close in size to Earth has been discovered in the cosmic vicinity of our Sun. The small size and proximity of the associated star bode well for studies of the planet's atmosphere. [SEE LETTER P.204](#)

## DRAKE DEMING

On page 204 of this issue, Berta-Thompson *et al.*<sup>1</sup> report the discovery of GJ 1132b, arguably the most important planet ever found outside the Solar System. The significance of this new world derives from several factors. It has a radius only 16% larger than Earth's and a matching density of 6 grams per cubic centimetre. Moreover, Gliese 1132,

the red dwarf star around which the planet orbits, lies only 12 parsecs from the Sun — a distance that will allow astronomers to study the planet with unprecedented fidelity.

The authors observe that GJ 1132b has an orbit that is nearly edge-on to an observer's line of sight, so it transits (passes in front of) Gliese 1132 once every 1.6 days. Although the planet is Earth-sized, the diameter of its star is only 21% that of the Sun (Fig. 1). Small stars



**Figure 1 | Relative sizes of planets to stars.** Berta-Thompson *et al.*<sup>1</sup> report an Earth-sized planet known as GJ 1132b, only 12 parsecs away from our Sun. All stars are large compared with Earth-like planets, but GJ 1132b's star (Gliese 1132) is considerably smaller than the Sun, which will facilitate studies of the exoplanet's atmosphere.

that host transiting exoplanets are invaluable, because their light interferes minimally with measurements of their planets' properties.

Exoplanet transits can yield a bonanza of scientific information, as NASA's Kepler space observatory has abundantly proved. By analysing transits, Kepler's discoveries include a planet that has two suns<sup>2</sup>, stars that host multiple planets<sup>3</sup> and a habitable super-Earth (a rocky planet several times more massive than Earth) that orbits a star similar to the Sun<sup>4</sup>. Beyond the romance of discovering new worlds, astronomers love transiting planets because the transit geometry allows them to unambiguously measure a planet's mass and radius (and thus to determine its bulk density), thereby providing basic information about its composition.

If a planet transits a bright, nearby star, the planetary atmosphere can also be probed. During a transit, a fraction of the star's light passes through the alien atmosphere. The atomic and molecular constituents of the atmosphere therefore become imprinted on the star's spectrum, which can be measured on Earth. The ultimate goal of such measurements is to identify biological 'signatures' in the atmospheres of nearby habitable worlds. But even if life is abundant in the cosmos, finding unambiguous spectroscopic evidence of it is a daunting task that lies in the indefinite future. Observations of small stars, such as Gliese 1132, loom large in this task.

Two intertwined quantities can be measured during an exoplanet transit. The first is stellar light transmitted through the planetary atmosphere, revealing the fingerprints of the planet's atmospheric composition. The second quantity is light from the star that misses the planet altogether — this is merely noise that unavoidably interferes with the planetary signal. Large stars are much bigger than their planets, and therefore produce a lot of this noise. By contrast, small stars more closely match the sizes of their planets, so the spectroscopic signals of the planets' atmospheres are preserved more clearly in transit spectra because there is less noise.

Moreover, small stars come with other advantages. These stars are abundant in our Galaxy, greatly outnumbering stars like the Sun (which, although sometimes described as an average star, is actually significantly larger than average). Because of this abundance, the closest habitable exoplanet to Earth will undoubtedly orbit a small star. More specifically, small red dwarfs such as Gliese 1132 emit a feeble output of light and heat compared with the Sun, and so any associated planets must orbit very close to their stars to stay warm enough to support life. That is a boon to astronomers searching for habitable planets, because close-in orbits have short durations, which means that planetary transits occur frequently — providing ample opportunities for the signature of life to be detected in atmospheric spectra.

Berta-Thompson and colleagues' results suggest that GJ 1132b is too warm to be habitable, but scientists have yet to fully explore our cosmic neighbourhood for worlds that potentially harbour life. The Kepler mission used a deep-distance search strategy to find as many planets as possible in a small fraction of the sky. But in 2017, NASA will launch the Transiting Exoplanet Survey Satellite<sup>5</sup> (TESS), which will use a complementary strategy: it will survey relatively nearby regions and cover the entire sky. TESS will find the habitable transiting planets that are closest to Earth, and thereby most amenable to study.

Shortly afterwards, a next-generation infrared observatory — the James Webb Space Telescope — will focus on the most amenable planets, including potentially habitable ones, to determine their atmospheric properties<sup>6</sup>. The atmospheres of such planets are likely to be different from Earth's, because the process of planet formation around small stars could differ distinctly from that of the Solar System, and, in the case of red dwarfs, because the spectrum of the star's light peaks at a much redder wavelength than does light from the Sun.

Although planets that transit small stars are good subjects for study in most respects, there are potential difficulties. The spectra

of red dwarfs similar to Gliese 1132 are rich in the signatures of molecules such as water, which astronomers are also hoping to detect in exoplanetary atmospheres. A precise knowledge of the spectrum of a star that illuminates a habitable exoplanet is therefore required to properly interpret the transit spectrum. Fortunately, work towards a detailed understanding of stellar spectra is progressing rapidly<sup>7</sup>, thanks to advances in infrared spectrometers that make the measurements, quantum chemistry that reveals the intrinsic properties of molecules, and stellar-atmospheric models that account for the complex structure of the outer layers of small stars. ■

**Drake Deming** is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA.  
e-mail: [ddeming@astro.umd.edu](mailto:ddeming@astro.umd.edu)

1. Berta-Thompson, Z. K. *et al. Nature* **527**, 204–207 (2015).
2. Doyle, L. R. *et al. Science* **333**, 1602–1606 (2011).
3. Lissauer, J. J. *et al. Nature* **470**, 53–58 (2011).
4. Jenkins, J. M. *et al. Astron. J.* **150**, 56 (2015).
5. Ricker, G. R. *et al. J. Astron. Telesc. Instrum. Syst.* **1**, 014003 (2015).
6. Beichman, C. *et al. Publ. Astron. Soc. Pacif.* **126**, 1134–1173 (2014).
7. Mann, A. W. *et al. Astrophys. J.* **804**, 64 (2015).

## CANCER

# The enemy of my enemy is my friend

**Reactive oxygen species have been viewed as stress-inducing molecules that promote cancer initiation. But new evidence indicates that oxidative stress can be beneficial — inhibiting the spread of a cancer to other sites. [SEE ARTICLE P.186](#)**

ISAAC S. HARRIS & JOAN S. BRUGGE

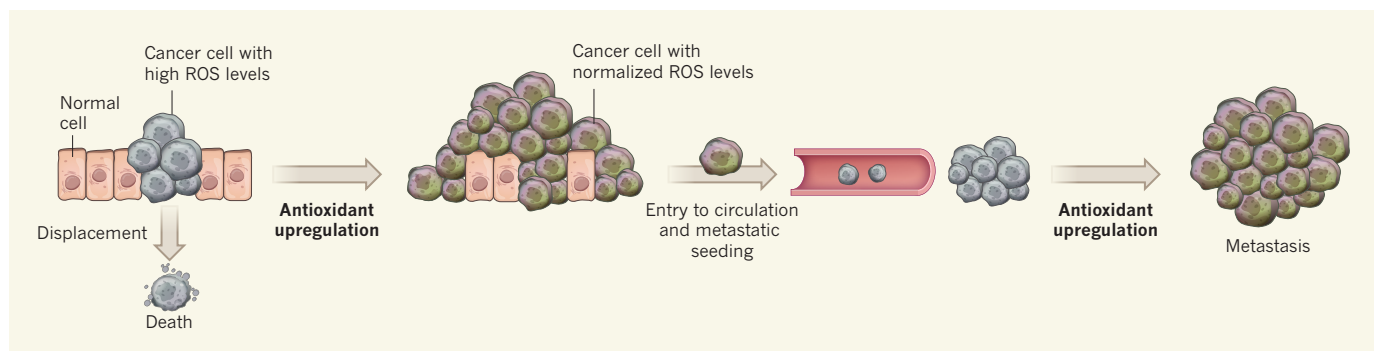
The roles of reactive oxygen species and the antioxidant molecules that detoxify them have been controversial in the field of cancer research because of their dichotomous ability to promote or suppress tumour development in different contexts<sup>1,2</sup>. In this issue, Piskounova *et al.*<sup>3</sup> (page 186) shed light on the subject by showing that efficient spread of melanoma cells to distant sites in the body depends on their ability to overcome the cellular stress induced by reactive oxygen species (ROS) that are encountered as the cells move through the bloodstream and initiate new tumours. The authors also demonstrate that inhibiting these metabolic pathways in mice effectively blunts the metastasis of melanomas.

ROS, which contain oxygen atoms with unpaired electrons, are highly reactive

molecules that damage DNA and other cellular components. Among the most abundant ROS are those generated when oxygen accepts free electrons to produce superoxide radicals ( $O_2^{\cdot-}$ ) and subsequently hydrogen peroxide ( $H_2O_2$ ). Most antioxidants are either enzymes that detoxify ROS or cofactors that participate in detoxification. Antioxidant cofactors can be synthesized through metabolic pathways, such as the one producing glutathione (GSH), or by protein synthesis, as for thioredoxin (TXN). A crucial partner in this process is NADPH, a cofactor generated by numerous metabolic pathways, which regenerates GSH and TXN after they participate in neutralizing ROS.

The oxidative stress exerted by ROS can lead to enhanced mutation rates, promoting the transformation of a normal cell into a tumour cell. ROS have also been shown to promote the stabilization of factors that can drive tumour





**Figure 1 | Metabolic adaptation to oxidative stress promotes metastasis.** As a tumour starts to grow, the elevated metabolic activity of the mutated cells, as well as uncontrolled proliferation that displaces cells from their normal environmental niches, can cause enhanced production of reactive oxygen species (ROS). These highly reactive molecules create a state of oxidative stress that can kill the cells; successful tumour cells combat this stress by producing antioxidant molecules that neutralize ROS<sup>9</sup>.

However, Piskounova *et al.*<sup>3</sup> provide evidence that tumour cells that detach from a primary tumour experience high ROS levels again, both in the bloodstream and at distant sites where they attempt to form metastatic tumours. The authors show that successful metastases are formed of tumour cells that have undergone metabolic changes that allow them to better withstand oxidative stress, including upregulating metabolic pathways that produce antioxidants.

initiation and progression<sup>4</sup>. Therefore, it was postulated that the use of antioxidants to quench ROS and reduce oxidative stress could represent both a preventive and a therapeutic anticancer strategy. On this basis, large-scale multicentre clinical trials of antioxidant supplementation were carried out<sup>5,6</sup>. However, this supplementation not only failed to benefit patients, but was associated with a significant increase in cancer incidence. Follow-up analysis in mouse models of cancer confirmed the pro-tumorigenic effects of antioxidant supplementation<sup>7</sup>. These outcomes posed the puzzling question: if antioxidants eliminate molecules that promote tumour initiation and progression, how can they also promote tumour growth?

Answers are beginning to emerge. During tumour initiation, cells undergo alterations — arising both from genetic aberrations and the influence of the microenvironment — that can increase the levels of ROS (Fig. 1). Excessive accumulation of ROS in the tumour cell can lead to growth arrest or death, thereby suppressing tumour progression. For example, aberrant hyperproliferation of cells leads to displacement from their natural 'niches' that provide essential growth factors; displaced cells are thus starved, leading to inefficient oxygen metabolism and generation of lethal levels of ROS<sup>8</sup>. Cellular drivers of tumour initiation and progression have been shown to increase production of antioxidants in an effort to combat the oxidative stress<sup>8,9</sup>, and dampening the production of these antioxidants can prevent malignant tumour initiation<sup>10</sup>. In addition, mutations that lead to the stabilization of NRF2, a protein that regulates antioxidant expression, have been found in multiple cancers, particularly lung cancer<sup>11,12</sup>. These findings help to explain how antioxidant supplementation can promote the expansion of tumours — by rescuing the viability of unhealthy, stressed tumour cells.

Piskounova *et al.* have now uncovered a role

for ROS that extends beyond tumour initiation to the inhibition of tumour metastasis. To identify the mechanisms underlying the differences in metastatic activity seen in melanomas, the authors carried out a systematic analysis of the metabolites present in cancerous cells following the transplant of human melanomas into mice. They found that tumour cells isolated from the circulation and from metastatic sites displayed higher levels of ROS than those from the primary tumour, and that the metastatic tumours showed a reversible elevation in the production of NADPH, which was associated with increased activity of the folate pathway, suggesting that the tumour cells underwent this adaptive change to buffer the oxidative stress<sup>13</sup>. Blocking the folate metabolic pathway, either by reducing expression of protein components of this pathway or by directly inhibiting its activity using the drug methotrexate, blunted melanoma metastasis in the mice. A role for ROS in melanoma metastasis was also reported recently in a study showing that administration of antioxidants strongly enhanced melanoma metastasis in mice<sup>14</sup>.

Numerous questions remain regarding the roles of antioxidants and ROS in cancer. For example, although Piskounova *et al.* highlight a role for the folate pathway in NADPH production, NADPH is also generated by other metabolic reactions, including those of the pentose-phosphate pathway. It will be useful to understand the contribution of these distinct pathways to NADPH generation in specific tumour contexts. Furthermore, it remains to be determined which antioxidants NADPH regenerates during the different stages of cancer initiation, progression and metastasis and in diverse tumour types.

A natural extrapolation of these findings is the potential use of antioxidant inhibitors to sensitize cancers to therapies by enhancing ROS-induced tumour-cell damage. However, such treatments would have to be fine-tuned to avoid oxidative damage to normal cells.

A greater understanding of such issues may lead to the development of tumour-specific antioxidant-targeted therapies.

Drugs that target antioxidants are currently being used to treat other diseases. For example, methotrexate, which Piskounova *et al.* used to target the folate pathway and inhibit antioxidant production, is a disease-modifying antirheumatic drug (DMARD). Sulfasalazine, which inhibits uptake of the amino-acid cystine, and auranofin, which prevents TXN regeneration, are also DMARDs. These drugs might provide a starting point for the identification of safe and effective inhibitors of antioxidants that could be explored for use in combination with standard cancer treatments. These findings, taken together, necessitate a revised perspective of the role of ROS in cancer, and force us to view it as an unlikely ally in our quest to treat this deadly disease. ■

**Isaac S. Harris and Joan S. Brugge** are in the Department of Cell Biology and the Ludwig Center at Harvard, Harvard Medical School, Boston, Massachusetts 02115, USA.  
e-mails: isaac\_harris@hms.harvard.edu;  
joan\_brugge@hms.harvard.edu

- Gorrini, C., Harris, I. S. & Mak, T. W. *Nature Rev. Drug Discov.* **12**, 931–947 (2013).
- Sabharwal, S. S. & Schumacker, P. T. *Nature Rev. Cancer* **14**, 709–721 (2014).
- Piskounova, E. *et al. Nature* **527**, 186–191 (2015).
- Gao, P. *et al. Cancer Cell* **12**, 230–238 (2007).
- Klein, E. A. *et al. J. Am. Med. Assoc.* **306**, 1549–1556 (2011).
- Chandel, N. S. & Tuveson, D. A. *N. Engl. J. Med.* **371**, 177–178 (2014).
- Sayin, V. I. *et al. Science Transl. Med.* **6**, 221ra15 (2014).
- Schafer, Z. T. *et al. Nature* **461**, 109–113 (2009).
- DeNicola, G. M. *et al. Nature* **475**, 106–109 (2011).
- Harris, I. S. *et al. Cancer Cell* **27**, 211–222 (2015).
- Kandoth, C. *et al. Nature* **502**, 333–339 (2013).
- Cancer Genome Atlas Research Network *Nature* **489**, 519–525 (2012).
- Fan, J. *et al. Nature* **510**, 298–302 (2014).
- Le Gal, K. *et al. Science Transl. Med.* **7**, 308re8 (2015).

This article was published online on 21 October 2015.

# Deadly combination

New evidence suggests that seismic waves from the Chicxulub meteorite impact doubled the eruption rate of lavas on the opposite side of the planet — a combination that led to the mass extinction at the end of the Cretaceous period.

ROBERT DUNCAN

The events surrounding the termination of the Cretaceous period 66 million years ago make one of the most captivating stories in our planet's history. Evidence for both a devastating meteorite impact at Chicxulub (Yucatan, Mexico) and massive volcanic eruptions in western India at this time is compelling, and each has been blamed for the abrupt demise of the dinosaurs, along with the majority of all species living at that time. A lively debate has raged for several decades about the predominance of one catastrophe or the other in the rapid and sustained environmental destruction that led to mass extinction. And now the story just gets better, with Renne and colleagues' report<sup>1</sup> in *Science* of evidence that the impact and the eruptions may be linked.

The Deccan Traps (Fig. 1) are an enormous, solidified accumulation of lava flows more than 10<sup>6</sup> cubic kilometres in volume<sup>2</sup>. They formed when melting of the mantle produced magmas that erupted through the continental lithosphere in western India. This volcanic activity is the most recent of several similarly large magmatic provinces whose occurrence coincides with mass extinctions through the geological record<sup>2</sup>. A limiting factor in understanding how such catastrophic volcanic activity leads to mass extinctions has been the uncertainty associated with

determining the age of the resulting rocks.

Previous ideas about the cause (or causes) of the terminal-Cretaceous mass extinction stemmed from the idea that species decline occurred gradually over several million years — which could be explained by a variety of factors, such as climate change, falling sea level and the advent of mammals that ate dinosaur eggs. Studies of high-resolution sedimentary records of the mass extinction<sup>3</sup>, however, have revealed that it was much more abrupt, leaving meteorite impact and massive volcanic eruptions as the only two viable mechanisms. Because Deccan volcanic activity started well before the Chicxulub impact<sup>4</sup>, fragments of the meteorite could not have initiated those eruptions, and so the impact and volcanism have been considered to be independent, coincidental events.

Renne *et al.* report high-resolution age determinations from the main section of Deccan lava flows (the Western Ghats), and so establish that the Chicxulub impact occurred within approximately 50,000 years of a dramatic change in eruption rate and magma composition. The researchers therefore assert that energy from the impact was propagated by means of seismic waves through and around Earth's surface, altering the Deccan magma 'plumbing system' to accelerate the eruption rate. This in turn contributed to environmental decline through the release of gases such as carbon dioxide and sulfur dioxide to the atmosphere.

The current findings follow on from recent related work by Richards *et al.*<sup>5</sup>, who estimated that the Chicxulub impact produced kinetic energy equivalent to a mega-earthquake of magnitude 9 to 11. They also reported that the seismic-wave energy arriving in the Deccan area, more than 14,000 km away from the impact, was sufficient to trigger volcanic eruptions by increasing the permeability of the crust and magma flow, and by causing the disruption or coalescence of crustal magma chambers<sup>6</sup>.

In the geological record, the Chicxulub impact is marked by high concentrations of the rare metal iridium and shocked quartz in marine sediments; shocked quartz forms when the normal form of the mineral is deformed under intense pressure but at low temperature at impact sites. The temporal coincidence of the impact with the mass extinctions at the end of the Cretaceous — known more formally as the Cretaceous–Palaeogene boundary (KPB) — is well accepted on the basis of those markers<sup>3</sup>, but the new data place the KPB much more precisely within the period of the Deccan lava flows.

The Chicxulub impact alone may have produced sufficiently severe and abrupt environmental deterioration to have caused mass extinction. By contrast, none of the other major mass extinctions shows evidence of meteorite impacts, but all coincide with the occurrence of large volcanic provinces<sup>2</sup>. It therefore seems reasonable to accept that both the impact and the volcanism, on nearly opposite sides of the planet, contributed to abrupt species decline and extinction at the KPB.

Richards *et al.*<sup>5</sup> also reported that eruption volume, frequency and magma composition during the Deccan lava-flow sequence changed substantially at about the time of the Chicxulub impact and the KPB. More than 70% of the total volume of magma erupted



LOÏC VANDERKLUYSEN

**Figure 1 | The Deccan Traps.** Massive lava flows in western India occurred at the end of the Cretaceous period, and produced the accumulation of rock known as the Deccan Traps. Renne *et al.*<sup>1</sup> suggest that the impact of the Chicxulub meteorite caused an acceleration in eruption rate during that volcanic activity, and that a combination of this increase and the effects of the impact may have caused the mass extinction that marked the end of the Cretaceous.



after this change, in the form of huge (up to 10,000 km<sup>3</sup>) lava flows during roughly the next 500,000 years<sup>7</sup>, although these flows occurred less frequently than before. But could this transition simply reflect thinning of the continental lithosphere, as documented for other large volcanic provinces<sup>8</sup>? Richards and colleagues note that the orientation of dikes — vertical, planar cracks in rock that allowed magma to rise to the surface — changed from orientations determined by the direction of lithospheric extension to randomly directed orientations across the transition, whereas trace-element indicators of depth of melting show no change. Both of these factors indicate that the increased magma production was not due to lithospheric extension and thinning.

Renne and colleagues' improved correlation of the Chicxulub impact with documented changes in eruption dynamics in the Deccan province supports a plausible mechanism for linking the former to the latter. This link does not fully explain the extended regime of larger but less-frequent eruptions following the impact, but the authors speculate that changes in magma-chamber size and distribution could be controlling factors. Ultimately, the KPB records a complex environmental response to two forcings — meteorite impact and volcanic eruptions — that are now firmly linked. Further details of the timescale for post-impact Deccan eruptions may emerge from the study of high-resolution marine sedimentary records of volcanic input and biotic recovery. ■

**Robert Duncan** is at the College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA.

e-mail: [rduncan@coas.oregonstate.edu](mailto:rduncan@coas.oregonstate.edu)

1. Renne, P. R. *et al. Science* **350**, 76–78 (2015).
2. Courtillot, V. E. & Renne, P. R. *C. R. Geosci.* **335**, 113–140 (2003).
3. Schulte, P. *et al. Science* **327**, 1214–1218 (2010).
4. Renne, P. R. *et al. Science* **339**, 684–687 (2013).
5. Richards, M. A. *et al. Geol. Soc. Am. Bull.* **127**, 1507–1520 (2015).
6. Manga, M. & Brodsky, E. *Annu. Rev. Earth Planet. Sci.* **34**, 263–291 (2006).
7. Chenet, A.-L. *et al. J. Geophys. Res.* **114**, B06103 (2009).
8. Fram, M. S. & Leshner, C. E. *Nature* **363**, 712–715 (1993).

## CLIMATE CHANGE

# A rewired food web

**Climate change is causing large fish species to move into arctic marine environments. A network analysis finds that these fishes, with their generalist diets, add links to the existing food web that may alter biodiversity and web stability.**

JULIA L. BLANCHARD

Climate change is driving species shifts both on land and in the sea. For species that inhabit extreme environments, such as the poles or highest altitudes, there may soon be nowhere left to go. The consequences of species redistributions for whole ecosystems are often discussed but poorly understood. This is due in part to the amount of information that is required to understand the complex networks of ecological interactions between organisms and their changing habitats. Writing in *Proceedings of the Royal Society B*, Kortsch *et al.*<sup>1</sup> combine network analysis with large-scale species-distribution data and detailed knowledge of food-web interactions — who eats whom — in the Barents Sea, which lies north of Norway and Russia. The authors use this information to show how changes in the distributions of the fastest-shifting fish species can substantially alter the structure of an arctic marine food web.

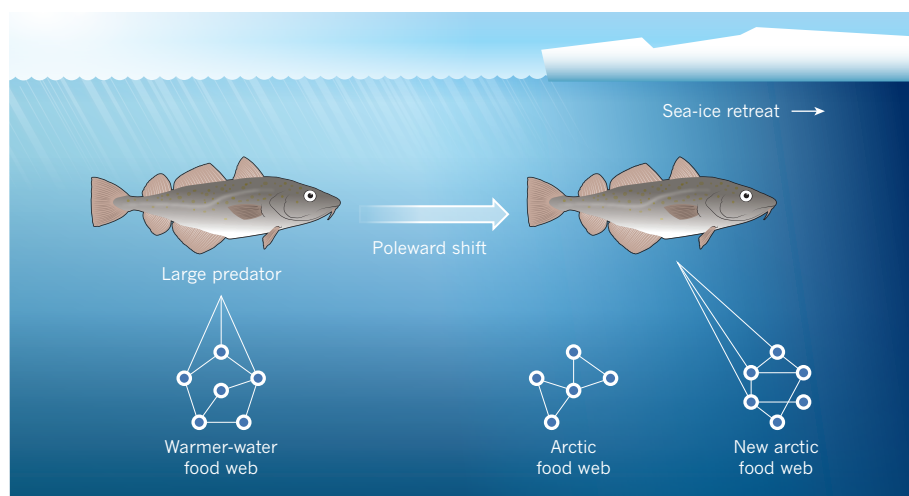
Warming waters mean that, in the Northern Hemisphere, some southerly fish species are moving northward and forging new interactions along the way. Will these changes push some species towards faster extinction and others towards domination? Although we do not yet have an answer, knowing the nature of the animals' interactions and how they are likely to change can provide clues to help guide the mitigation of biodiversity loss before it is too late.

The study of complex networks has helped us to understand how quickly diseases can spread, how information moves on the Internet, and

how traffic, trade and even biochemical signals work<sup>2</sup>. Food webs are complex networks of predator–prey interactions that have fascinated scientists for more than a century. The structure of a food web — how the links between species are configured — can influence the web's stability. A variety of metrics describe network connections. These include the number of links to and from each species; the strength and density of the links; whether subsets of the web are clustered in modules; and whether there are

loops within networks. Such metrics have been studied for their influence on multiple ecosystem features<sup>3</sup>, such as whether simulated species extinctions lead to cascades of other species going extinct, or how fast disturbances spread through the network.

Kortsch and colleagues' network analysis of marine arctic food webs shows that they contain relatively few connections and are quite modular — two features that have been linked to greater stability and resilience to the spread of perturbations. These food webs function in regions that are covered by sea ice for a large proportion of the year. Here, the most abundant life forms are benthic organisms, which live on or beneath the sea floor, and the food they eat comes from dead or dying tiny algae (phytoplankton) that are normally locked in the sea ice. This benthic energy flow has been described as a slow pathway, one that is relatively resilient to sudden or abrupt changes<sup>4,5</sup>.



**Figure 1 | Shifting structure.** The marine arctic food webs studied by Kortsch *et al.*<sup>1</sup> contained fewer links than those farther south. This is partly because the warmer waters are home to large, wide-roaming fish species that consume organisms of a range of sizes and in various habitats. However, these large fishes are starting to move poleward as climate change causes warming and sea-ice retreat. The authors document that the increasing presence of large fishes in marine arctic environments is changing the structure of the food web, introducing new links and decreasing its modularity.

By contrast, farther south, in the warmer and ice-free subarctic region, marine food webs contain many more links, including those of large fish species that are wide-ranging, are present throughout the year and are generalists (they are not picky eaters). In part because of their size, these fish species act as super connectors, and they have a vast menu from which to choose, consuming organisms of a wide range of sizes — not only species that occupy the sea floor, but also those in the entire water column (the pelagic zone). Through predation, these large fish often keep smaller predators and grazers in check.

How will arctic food webs change if species move north? Four species of large generalist fish are shifting rapidly poleward<sup>6</sup>: cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*) and two redfish species (*Sebastes norvegicus* and *Sebastes mentella*). Using knowledge of what these fish eat<sup>7</sup> and detailed geographical information on current species locations, Kortsch *et al.* added these species and their links to the marine arctic food web and identified a shift in the web's structure (Fig. 1).

What they find is not surprising — compared with the original web, the revised web has more links and lower modularity, essentially making it more similar to food webs farther south. Surprises may come, however, in how these changes manifest themselves throughout the ecosystem over time. The web structure alone does not allow us to predict changes in species abundance, but it does provide a 'road map' with which to investigate how potential changes could spread through the network. For example, loops in food webs can cause counterintuitive or unexpected outcomes. The original arctic food web did not contain loops, but the new web does.

Predicting how these various changes will affect ecosystem stability is tricky. With lower modularity and more links from large generalists, the revised web may become more susceptible to perturbations (because these spread faster) or more robust (through regulation by predators). The outcome also depends on the type of perturbation and the characteristics of the individuals present, such as body size and physiological performance.

As the sea ice retreats, it may open up greater access for human activities, such as fisheries and oil exploration. As well as being large generalists, the shifting fish species are all commercially important. Food-web interactions form the core of ecosystem models that are commonly used to assess the impacts of human activity on ecosystems or non-target species. However, the food-web linkages in such models are usually assumed to remain constant. Kortsch and colleagues' study contributes to the mounting evidence that these structures change over time<sup>8</sup>.

Climate change is also causing other types of ecosystem change. Abrupt transitions

called regime shifts are affecting the balance of macroalgae and benthic invertebrates in the same region<sup>9</sup>, mirroring patterns that have been reported globally<sup>10</sup>. Given the inevitability of change for these potentially vulnerable ecosystems, we will require a better dynamical understanding of food-web rewiring and how it interacts with multiple traits and stressors. This insight should be integrated with cumulative assessments of impacts on the ecosystem, to minimize the collateral damage that climate change and human activities might inflict on biodiversity in the future. ■

**Julia L. Blanchard** is at the *Institute of Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania 7001, Australia.*  
e-mail: julia.blanchard@utas.edu.au

#### MATERIALS CHEMISTRY

## Liquefied molecular holes

**Porous solids have many uses in the chemical industry, which has stimulated the development of several generations of such materials. A new generation has now arrived, with the report of permanently porous liquids. SEE LETTER P.216**

**MICHAEL MASTALERZ**

**A**lthough many kinds of porous solid are known, permanently porous liquids are rare — most liquids contain only transiently formed cavities between their molecules. But on page 216 of this issue, Giri *et al.*<sup>1</sup> report a trick that enables permanently porous liquids to be formed from organic cage-like molecules. Their compounds represent a new class of material that might one day be used in applications such as trapping carbon dioxide emitted from industrial sites, for which a liquid capture system has practical advantages.

Porosity is a common natural phenomenon often found in living systems, including tree bark and marine sponges. Pores have also been found in certain inorganic minerals such as stilbite<sup>2</sup>. Heating these minerals causes steam to be released, because the pores contain large amounts of water. This class of mineral was therefore named zeolites — a word derived from the greek *zeo* and *lithos*, meaning boiling stone.

Zeolites are chemically and thermally robust, and have found several applications, especially in the chemical industry. They are particularly useful as catalysts because their surface areas are larger than those of similar non-porous compounds, making them more catalytically active. But only a limited number of zeolites are available.

1. Kortsch, S., Primicerio, R., Fossheim, M., Dolgov, A. V. & Aschan, M. *Proc. R. Soc. B* **282**, 20151546 (2015).
2. Dunne, J. A., Williams, R. J. & Martinez, N. D. *Proc. Natl Acad. Sci. USA* **99**, 12917–12922 (2002).
3. Thompson, R. M. *et al. Trends Ecol. Evol.* **27**, 689–697 (2012).
4. Rooney, N., McCann, K., Gellner, G. & Moore, J. C. *Nature* **442**, 265–269 (2006).
5. Blanchard, J. L., Law, R., Castrillo, M. D. & Jennings, S. *Theor. Ecol.* **4**, 289–300 (2011).
6. Fossheim, M. *et al. Nature Clim. Change* **5**, 673–677 (2015).
7. Planque, B. *et al. Ecology* **95**, 1430 (2014).
8. Gilljam, D., Curtsdotter, A. & Ebenman, B. *Nature Commun.* **6**, 8412 (2015).
9. Kortsch, S. *et al. Proc. Natl Acad. Sci. USA* **109**, 14052–14057 (2012).
10. Ling, S. D. *et al. Phil. Trans. R. Soc. B* **370**, 20130269 (2015).

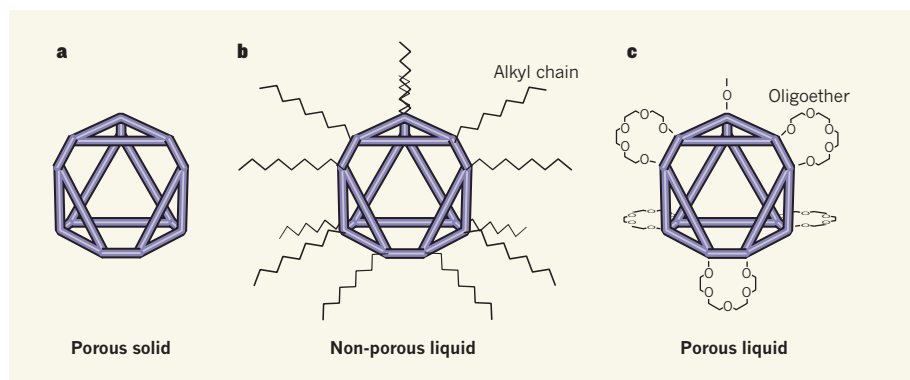
This article was published online on 4 November 2015.

This problem was solved by the advent of porous solids called metal–organic frameworks (MOFs)<sup>3</sup>. In these materials, metal ions or charged metal–oxygen clusters act as the nodes of 2D or 3D molecular frameworks, with organic linker molecules as the connecting struts. This construction principle enables porous networks to be designed, and has been hugely influential — a vast number of porous solids has become available in the past few decades<sup>4</sup> using this or related synthetic strategies. Some of these materials have large surface areas and distinct properties, such as a high adsorption capacity for CO<sub>2</sub>.

In 2009, another branch in the evolutionary tree of porous materials appeared: shape-persistent molecular cages<sup>5</sup>. In these systems, organic cage-like molecules form crystalline materials that retain the void space of the cages. An advantage of molecular cages is that they can be dissolved without any of the chemical bonds disconnecting. Such cages can therefore be thought of as soluble porous units, which means that they can be processed for incorporation into other materials or to make thin-film devices<sup>6,7</sup>.

At ambient temperature, liquids are usually composed of molecules. Organic cage molecules could therefore in principle be used to make porous organic liquids<sup>8</sup>. Liquids are desirable, because they are easier to transport than solids (because they can be pumped





**Figure 1 | How to make permanently porous liquids.** **a**, Organic cage molecules pack together to form crystalline porous solids. **b**, When long alkyl chains are attached to the cages, the number of packing interactions decreases, and a solid cannot form at 50 °C. However, the resulting liquid is not porous because the chains penetrate the cages' cavities. **c**, Giri *et al.*<sup>1</sup> report that cages with attached cyclic oligoether units form a permanently porous liquid when mixed with a solvent (15-crown-5), because neither the solvent nor the oligoethers can enter the cage cavities.

through tubes) and can be easier to process — for instance, they can be ‘painted’ onto surfaces to make thin films. So what are the molecular requirements for a compound to be liquid at ambient conditions?

The main requirement is to minimize the weak intermolecular forces that pack molecules together. Organic chemists usually do this by attaching long, sometimes branched, hydrocarbon (alkyl) chains to molecules that normally pack together. Researchers from Giri and colleagues' laboratory previously reported<sup>9</sup> that the melting points of organic cages are indeed substantially lowered by the attachment of long alkyl chains (Fig. 1). Unfortunately, it turned out that these liquids are not porous, possibly because the alkyl chains penetrate the cage cavities and block the pores.

Giri *et al.* now report the successful production of permanently porous liquids using a simple trick — their cage molecules still have flexible chains attached, but interpenetration is avoided by attaching looped chains instead of linear chains. This was achieved using oligoether units (small hydrocarbon chains that are connected by oxygen atoms). The resulting compound has a melting point higher than 180 °C, and so is not a porous liquid by itself, but the authors created a porous liquid by combining the cages with a solvent called 15-crown-5 (the ratio of cage to solvent was 1:12, and probably corresponds to the highest concentration of cages that could be achieved in 15-crown-5). The molecules of this solvent are too big to enter the cages' cavities.

The 15-crown-5 solvent has a major role in maintaining the cage as part of a flowing liquid at ambient conditions. Other solvents that are too bulky to enter the cages could also be used, and the oligoether groups can be replaced by smaller groups that cannot penetrate the cavities. However, the solubilities of cage compounds that have small peripheral groups are, in general, relatively low in common

organic solvents, and so the number of pores that they would create in the liquid is also relatively low.

Giri *et al.* demonstrate that cages with small peripheral groups can also be used to create porous liquids by making scrambled cages — a blend of cage compounds in which the six oligoether groups are replaced by a statistical mixture of pairs of methyl groups and six-membered hydrocarbon rings (see Fig. 4a of the paper<sup>1</sup>). The solubility of the scrambled cages in hexachloropropene (another solvent that is too large to enter the cavities) is greater than those of cages that contain only methyl pairs or six-membered rings as peripheral groups. The resulting liquid has a similar porosity to that of the material generated from oligoether cages and 15-crown-5, but is much less viscous. For both porous liquids, obtaining the correct combination of molecular structure and solvent is essential for success. Because the scrambled cages are easier to prepare than the oligoether cages, the porous liquid made from the scrambled cages might have more promise for future applications.

The researchers proved that their liquids are porous using a sophisticated spectroscopic technique, but also used the naked eye. When gas was adsorbed in a porous liquid and a solvent small enough to enter the cages' cavities was added, displacement of the gas by solvent was immediately detectable by the evolution of bubbles in the liquid.

The surface area per unit volume and the overall uptake of gases in the porous liquids are much lower than those of solid porous materials, so the liquids cannot compete for applications immediately — they should instead be seen as a prototype of a new class of material. But if the amount of gas that can be adsorbed within these liquids can be increased, then this new generation of porous materials will undoubtedly find technological applications, such as in liquid porous beds for efficient gas separations or gas chromatography<sup>10</sup>. ■



## 50 Years Ago

Mr. K. E. B. Jay, who was well known for his popular books on atomic energy, died at his home in East Hendred, Berkshire, on August 3, at the age of fifty-five. ... In all his writing Jay took immense pains both to keep in mind the needs of his readers for a clear and simple exposition, and also to preserve scientific accuracy. In this way he was able to achieve the objectives he set himself, and at the same time to retain fully the confidence of the scientists about whose work he wrote: nearly all his writing was based on first-hand discussion with them. ... He was an excellent lecturer both in the history of atomic energy and on the presentation of scientific information ... Science is becoming more complex, more expensive and more difficult to understand; at the same time, its understanding by laymen in industry and Government must increase if science is to be efficiently applied. Kenneth Jay had an outstanding ability to bridge this gap in communication between scientists and laymen; there is a great need for many more like him.

From *Nature* 13 November 1965

## 100 Years Ago

A further appeal is made for the presentation or loan of telescopes for use with the batteries at the Front. Large telescopes on stands, deer-stalking telescopes, good pocket telescopes — in short, any type except toys are acceptable. Astronomical eye-pieces, etc., not wanted on service, are removed before issuing and marked with a number identical with that engraved on the telescope. Offers or instruments should be sent to the secretary, Lady Robert's Field Glass Fund, National Service League, 72 Victoria Street, S.W.

From *Nature* 11 November 1915

Michael Mastalerz is at the Institute for Organic Chemistry, Ruprecht-Karls-Universität Heidelberg, 69120 Heidelberg, Germany.  
e-mail: michael.mastalerz@oci.uni-heidelberg.de

1. Giri, N. *et al. Nature* **527**, 216–220 (2015).
2. Barrer, R. M. *Zeolite and Clay Minerals as Sorbents and Molecular Sieves* (Academic, 1978).
3. Yaghi, O. M., Li, G. & Li, H. *Nature* **378**, 703–706 (1995).
4. Slater, A. G. & Cooper, A. I. *Science* **348**, aaa8075 (2015).
5. Tozawa, T. *et al. Nature Mater.* **8**, 973–979 (2009).

6. Hasell, T., Zhang, H. & Cooper, A. I. *Adv. Mater.* **24**, 5732–5737 (2012).
7. Brutschy, M., Schneider, M. W., Mastalerz, M. & Waldvogel, S. R. *Adv. Mater.* **24**, 6049–6052 (2012).
8. O'Reilly, N., Giri, N. & James, S. L. *Chem. Eur. J.* **13**, 3020–3025 (2007).
9. Giri, N. *et al. Chem. Sci.* **3**, 2153–2157 (2012).
10. Kewley, A. *et al. Chem. Mater.* **27**, 3207–3210 (2015).

## BIOLOGICAL RHYTHMS

# Human sleep before the industrial era

**A quantitative study of sleep patterns in three pre-industrial societies implies that our natural sleep duration is close to seven hours, and that sleep cycles are determined by environmental temperature as well as the light–dark cycle.**

DERK-JAN DIJK & ANNE C. SKELDON

How much sleep do we need? What time should we start school or work so that we are in tune with our body clocks? These are topical questions, and hardly a week goes by without media reports about the importance of sleep and daily rhythms for physical and mental health. The suggestion is that our modern sleep patterns are muddled. But what are our natural sleep patterns, and how are they influenced by environmental factors? What is the benchmark for 'good' sleep? Writing in *Current Biology*, a team of sleep neuroscientists and anthropologists (Yetish *et al.*<sup>1</sup>) reports a study of sleep patterns in three hunter-gatherer and hunter-horticulturalist groups. During the study, and as is usual for these communities, none of the participants had access to electricity, and

the only sources of light, apart from the Sun and Moon, were small fires.

The study participants came from the Hadza group in Tanzania, the Ju/'hoansi San in Namibia and the Tsimane' in Bolivia. The authors monitored the participants' sleep over several days or weeks using activity and light recorders, and found that the average sleep duration across the three groups — the interval between sleep onset and sleep end — was 7.7 hours (Fig. 1). Subtracting the awakenings that occur during the night left an average total sleep time of only 6.4 hours. How does this compare with sleep in modern industrialized societies that use electricity daily? The embarrassing answer is: we don't really know, because most estimates of sleep duration are based on self reports rather than the objective measures used by Yetish and colleagues. However, the self-reported average sleep duration

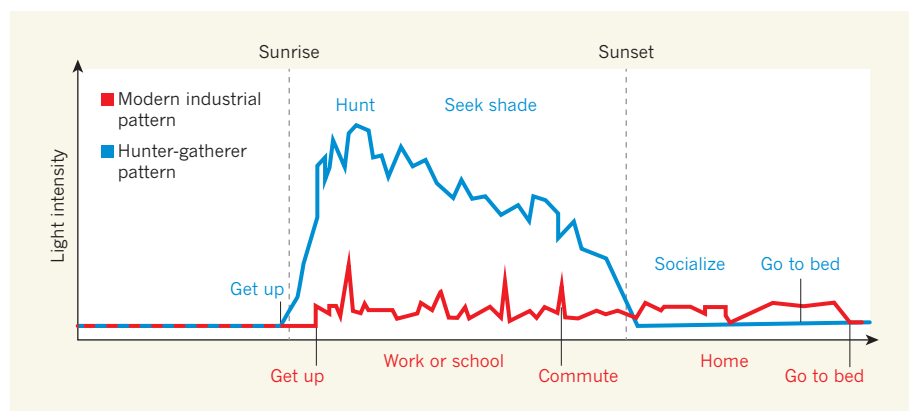
in modern industrialized groups is around 7–7.5 hours, and although it depends on the day of the week and the age of the individual<sup>2,3</sup>, it differs little from that found in this study.

When do the 7.7 hours of sleep recorded by Yetish *et al.* occur? No watches or electronic clocks were available to the participants, so sleep timing was determined by environmental cues and internal biological clocks. Maybe surprisingly for those of us who think that napping is natural, the authors found that few daytime naps were taken. Furthermore, sleep did not begin at the onset of darkness, but on average 3.3 hours after sunset. This average conceals a remarkable night-to-night variation in sleep onset — it seems that a regular bedtime is not a hallmark of natural sleep. By contrast, individuals in each group tended to wake at similar times, generally before sunrise. How does this compare with modern, industrialized sleep? Many people do go to sleep well after sunset, but few of us consistently rise before sunrise.

The timing of natural sleep can be easily linked to what we know about our biological daily rhythms. Earth's rotation results in environmental cycles of light and dark and hot and cold, and evolution has favoured the survival of biological mechanisms that predict these daily geophysical patterns. This rhythmicity has been observed right down to the cellular and molecular level, with nearly every cell in the human body showing approximately 24-hour (circadian) oscillations in gene expression. It is thought that synchrony of these billions of individual cellular rhythms is orchestrated by the brain's master clock, and it is well established that the light–dark cycle is the most prominent environmental stimulus for synchronizing this clock with the external world. The timing of our internal clock determines when we feel sleepy.

However, since the discovery of fire, we have learnt to manipulate our exposure to light. We can extend the light period by lighting a fire, a candle, an oil lamp, an incandescent light bulb, a fluorescent lamp or a light-emitting diode, giving us behavioural control over the stimulus that synchronizes our clock. The participants in Yetish and colleagues' study have this control to only a limited extent — the dim red light emitted by their fires has less biological effect than electric light with a strong blue component, such as is emitted by electronic devices and many low-energy light bulbs<sup>4,5</sup>.

Recent studies exploring the impact of introducing electric light to the Toba Qom people in the Argentinean Chaco region<sup>6</sup>



**Figure 1 | Daily patterns.** Yetish and colleagues' study<sup>1</sup> of sleep patterns in hunter-gatherer and hunter-horticulturalist groups allows a comparison between typical sleep patterns and daily activity schedules of individuals in modern but pre-industrial societies and those in modern, industrial societies<sup>4</sup>. In the latter societies, regular access to electricity divorces people from the natural environmental light–dark cycle, with light-exposure levels being largely self-determined and fairly uniform across waking hours. The authors' findings show that average sleep duration in the two societies is similar, but that sleep timing is more in sync with environmental cycles in pre-industrial than in industrial societies. (Schematic of light intensity derived from refs 1 and 4.)



or to rubber tappers in the Amazon<sup>7</sup> indeed demonstrate delayed bedtimes and reduced sleep duration. And evidence from a study comparing US residents in daily life and on a camping trip away from artificial light suggests that access to electric light not only shifts the circadian clock but also exaggerates the natural individual variability of sleep timing<sup>8</sup>. At an extreme, perhaps the owl-like tendencies of adolescents, and the resulting debate on shifting school timing to later starts, are to a considerable extent a direct result of our manipulation of our light environment.

Yetish and colleagues also point to the influence of a second environmental cycle on sleep timing: temperature. Core body temperature rises and falls over each 24 hours, with sleep generally occurring as the body temperature falls. Because we are a diurnal species, awake during the day and asleep during the night, and because environmental temperature tends to drop during the hours of darkness, the rhythm of our core body temperature tends to align with that of the environmental temperature. From an energetic perspective, this makes sense — keep the difference between environment and body temperature as small as possible and you reduce the amount of energy needed to stay warm. This link between environmental temperature, metabolic demands and the timing of the sleep–wake cycle has received attention in the animal and human sleep literature<sup>9,10</sup>. Yetish and colleagues found that the timing of waking is closely associated with the environmental minimum temperature; in the one group that woke after sunrise, the San, this pattern occurred only

during summer, when the environmental temperature minimum also occurred after sunrise.

We are beginning to understand the impact of our artificial world on our sleep–wake rhythms. Although we have tantalizing hints about how the light environment of our modern world affects our sleep patterns, there are few data on how we manipulate our temperature environment and the effects of such manipulation on sleep. Yetish *et al.* have provided findings that alter our assumptions about sleep in our ancestors, and have opened the door to further studies of the effects of light and temperature on sleep today. ■

**Derk-Jan Dijk** is in the Surrey Sleep Research Centre, University of Surrey, and **Anne C. Skeldon** is in the Department of Mathematics, University of Surrey, Guildford GU2 7XP, UK.  
e-mails: d.j.dijk@surrey.ac.uk;  
a.skeldon@surrey.ac.uk

1. Yetish, G. *et al.* *Curr. Biol.* <http://dx.doi.org/10.1016/j.cub.2015.09.046> (2015).
2. Roenneberg, T. *Nature* **498**, 427–428 (2013).
3. Skeldon, A. C., Derks, G. & Dijk, D.-J. *Sleep Med. Rev.* <http://dx.doi.org/10.1016/j.smr.2015.05.011> (2015).
4. Santhi, N. *et al.* *J. Pineal Res.* **53**, 47–59 (2012).
5. Czeisler, C. A. *Nature* **497**, S13 (2013).
6. de la Iglesia, H. O. *et al.* *J. Biol. Rhythms* **30**, 342–350 (2015).
7. Moreno, C. R. C. *et al.* *Sci. Rep.* **5**, 14074 (2015).
8. Wright, K. P. Jr *et al.* *Curr. Biol.* **23**, 1554–1558 (2013).
9. van der Vinne, V. *et al.* *Proc. Natl Acad. Sci. USA* **111**, 15256–15260 (2014).
10. Daan, S., Honma, S. & Honma, K. *J. Biol. Rhythms* **28**, 403–411 (2013).

## ECONOMICS

# Higher costs of climate change

**An attempt to reconcile the effects of temperature on economic productivity at the micro and macro levels produces predictions of global economic losses due to climate change that are much higher than previous estimates. [SEE LETTER P.235](#)**

THOMAS STERNER

**W**e are already experiencing the economic impacts of climate change — heatwaves, for example, are increasing health costs and employee absenteeism, as well as reducing crop yields. But attempts to calculate the costs of warmer temperatures have produced conflicting results, particularly between estimates at the micro versus the macro scale in wealthy countries. Aggregating cost estimates from many different instances of micro-scale damage to obtain

a single macro-scale estimate for the whole economy is very hard. In this issue, Burke *et al.*<sup>1</sup> (page 235) show that these inconsistencies can be reconciled if nonlinearity in the relationship between temperature and economic productivity is taken into account at the macro scale. Furthermore, their results imply that the damages from climate change are much more serious than is generally believed.

If a cyclone hits your house, the correct cost of the damage is not what the house originally cost but the cost of the best replacement you can make that will leave you equally well



**Figure 1 | Dry fields in Shanxi province, northern China.** Many micro-level economic outputs, such as crop yields, show abrupt, nonlinear or even discontinuous relationships with temperature, decreasing dramatically at high temperatures. Burke *et al.*<sup>1</sup> find that a smooth but nonlinear response to temperature also applies to macro-level economic productivity in both wealthy and poor countries.

off. Now extrapolate this example to climate change. If you realize that your house was badly designed or badly located to withstand a changing climate, then the best replacement might be to rebuild the house somewhere else or even to spend the money in an entirely different way. This example shows how damage motivates adaptation. But adaptation can, in turn, change the cost of damage. Thus, in the aftermath of major events related to climate change, we recalculate the future course of action and reallocate resources. This is one of the reasons why simply looking at the micro-level costs of any disastrous event gives an incomplete picture, and why macro-level evidence is needed.

Burke and colleagues set themselves the task of connecting micro- and macro-level estimates of the costs related to changes in temperature and other climate variables. A small, but increasing, number of studies have shown that various micro-level components of the economy exhibit a highly nonlinear response to local temperature in a wide variety of countries, both rich and poor (see, for instance, refs 2 and 3). For example, worker productivity and crop yields are both relatively stable at temperatures between 0 and 25 °C, but decline steeply at higher temperatures (see Fig. 1 of the paper<sup>1</sup>). The question is how to aggregate such effects to cover the whole economy without double counting or missing vital parts. One seminal study<sup>4</sup> found no correlation between macroeconomic productivity and temperature in rich countries, but a linear correlation in low-income countries (that is, the higher the temperature, the bigger the costs).

Burke *et al.* find several differences from that study. They analysed updated and slightly

different data covering several additional years (2004–10). They also used a different (and I believe improved) approach to handling confounding variables in their models, together with other methodological details that give somewhat higher precision in their estimates. The main result is an overall nonlinear pattern in the relationship between temperature and economic growth. Almost all low-income countries are in 'warm' regions, and thus are predicted to suffer strong effects when temperatures go even higher, whereas rich industrialized countries are typically closer to the 'optimal' average temperature and thus show a weaker and more varied response. There are many possible ways of running such regression analyses, but the authors chose to focus on estimates that do not compare one country with another (which would risk inviting the influence of many confounding factors), but instead compare each country with itself during years with different temperatures.

The authors also sought to reconcile their aggregate finding with evidence from micro-level economic activities that exhibit different temperature dependencies. Although these relationships vary and are typically strongly nonlinear, aggregation of these effects can smooth the pattern, resulting in a nonlinear macro-level response to temperature that is essentially applicable to all countries (Fig. 1). According to the authors' modelling, overall economic productivity peaks at an annual average temperature of 13 °C and declines strongly at higher (and lower) temperatures. This relationship seems to apply to all countries, to be constant since 1960 and to be applicable to both agricultural and non-agricultural activities. The implication is that all kinds of economic activity

in all kinds of countries are heavily influenced by changes in our global climate.

The authors take great care to check the robustness of their findings but there will, no doubt, be attempts to look for other data and approaches, which may give different results. But such is the scientific process and, should these conclusions stand up or even be strengthened, they will have far-reaching implications. For example, unmitigated warming is predicted to lead to a (population weighted) average temperature increase of 4.3 °C by 2100: Burke and colleagues' model predicts that the accompanying climate changes would lead to a decline in average global incomes by a quarter compared with a scenario without climate change. Global income would also become more unequal, because some regions would benefit (particularly those that have a cold climate today, which tend to be countries with quite high incomes) whereas others, especially the (warmest and) poorest, would be hit very hard.

All told, these estimates equate to much larger economic losses than most leading models suggest (assuming that we do not find completely new ways to adapt), and hence give even more reason to mitigate damages today. The current leading models, referred to as integrated assessment models (IAMs), are already being used as a basis for policy. In the United States, there have been considerable battles, even in Congress, concerning the 'social cost of carbon', which is based on the three most prominent IAMs (see, for example ref. 5). Burke and colleagues' results suggest that these damage predictions, and thus also the social cost of carbon, need to be raised by several hundred per cent.

The conclusion that temperature-associated costs will be higher than previously calculated will cause a stir, and should have stark repercussions for policy. Another major significance of this paper will be to expand this emerging field of work. Future assessments of the relationship between economics and temperature will inevitably change details and expose nuances, particularly concerning the role of adaptation. My feeling is that we are only beginning to understand just how much damage a changed climate can wreak. ■

**Thomas Sterner** is in the Department of Economics, University of Gothenburg, SE 405 30 Gothenburg, Sweden.  
e-mail: [thomas.sterner@economics.gu.se](mailto:thomas.sterner@economics.gu.se)

1. Burke, M., Hsiang, S. M. & Miguel, E. *Nature* **527**, 235–239 (2015).
2. Schlenker, W. & Roberts, M. J. *Proc. Natl Acad. Sci. USA* **106**, 15594–15598 (2009).
3. Graff Zivin, J. & Neidell, M. J. *Labor Econ.* **32**, 1–26 (2014).
4. Dell, M., Jones, B. F. & Olken, B. A. *Am. Econ. J.: Macro* **4**, 66–95 (2012).
5. Revesz, R. L. *et al. Nature* **508**, 173–175 (2014).

This article was published online on 21 October 2015.



# Basomedial amygdala mediates top-down control of anxiety and fear

Avishek Adhikari<sup>1,2\*</sup>, Talia N. Lerner<sup>1,2\*</sup>, Joel Finkelstein<sup>1\*</sup>, Sally Pak<sup>1</sup>, Joshua H. Jennings<sup>1,2</sup>, Thomas J. Davidson<sup>1,2</sup>, Emily Ferenczi<sup>1,3</sup>, Lisa A. Gunaydin<sup>1,3</sup>, Julie J. Mirzabekov<sup>1</sup>, Li Ye<sup>1,2</sup>, Sung-Yon Kim<sup>1,3</sup>, Anna Lei<sup>1</sup> & Karl Deisseroth<sup>1,2,3,4,5</sup>

**Anxiety-related conditions are among the most difficult neuropsychiatric diseases to treat pharmacologically, but respond to cognitive therapies. There has therefore been interest in identifying relevant top-down pathways from cognitive control regions in medial prefrontal cortex (mPFC). Identification of such pathways could contribute to our understanding of the cognitive regulation of affect, and provide pathways for intervention. Previous studies have suggested that dorsal and ventral mPFC subregions exert opposing effects on fear, as do subregions of other structures. However, precise causal targets for top-down connections among these diverse possibilities have not been established. Here we show that the basomedial amygdala (BMA) represents the major target of ventral mPFC in amygdala in mice. Moreover, BMA neurons differentiate safe and aversive environments, and BMA activation decreases fear-related freezing and high-anxiety states. Lastly, we show that the ventral mPFC–BMA projection implements top-down control of anxiety state and learned freezing, both at baseline and in stress-induced anxiety, defining a broadly relevant new top-down behavioural regulation pathway.**

Human<sup>1,2</sup> and animal<sup>3,4</sup> studies have implicated diverse cortical and subcortical regions in anxiety and fear regulation. Interestingly, altered structure<sup>1</sup> and activity correlations<sup>2</sup> between mPFC and amygdala have been reported in patients with anxiety disorders, although the precise causal connections remain unclear<sup>3,5</sup>. Complexity is suspected, since ventral and dorsal mPFC (vmPFC and dmPFC, respectively) may have opposing roles in fear (vmPFC inhibits dmPFC<sup>6</sup>, and stimulation of vmPFC<sup>7</sup> or dmPFC<sup>3</sup> respectively decreases or increases freezing). Relevant subcortical regions are also complex; inhibitory intercalated cells (ITCs) in amygdala have been hypothesized to be vmPFC targets<sup>8</sup>, and to inhibit fear-promoting cells of the central nucleus of the amygdala, which could be relevant to the decreased freezing caused by electrical stimulation of vmPFC<sup>3</sup>. In contrast, dmPFC innervates the basolateral amygdala (BLA)<sup>9</sup>, and the bulk of the BLA population promotes fear<sup>10–12</sup>. This model could explain vmPFC–dmPFC functional differences<sup>3,10</sup> and why lesioning ITCs promotes freezing<sup>13</sup>, but has never directly and precisely been tested.

In rats the vmPFC–ITC projection is sparse<sup>8</sup>, suggesting consideration of other targets of mPFC in amygdala for mediation of top-down control. Moreover, prior investigations of mPFC–amygdala circuitry employed electrical<sup>13–15</sup> or optical stimulation that did not precisely resolve projections defined by cortical subregion origin and amygdala subregion target<sup>7,16,17</sup>. Finally, slower-timescale vmPFC<sup>18–21</sup> and dmPFC<sup>18,22,23</sup> lesions and inactivations have generated conflicting effects on anxiety. The precise identity of a functional top-down anxiolytic circuit has thus remained unknown. Here, we resolved distinct mPFC–amygdala projections in mice by combining anatomical tracing, CLARITY, *in vivo* and *in vitro* mapping of functional connectivity, and optogenetic control of mPFC–amygdala projections during fear- and anxiety-related behaviours.

## Direct top-down control of anxiety

Increased anxiety involves alterations in physiology and behaviour (for example, increases in respiratory rate and risk-avoidance)<sup>24</sup>. In rodents,

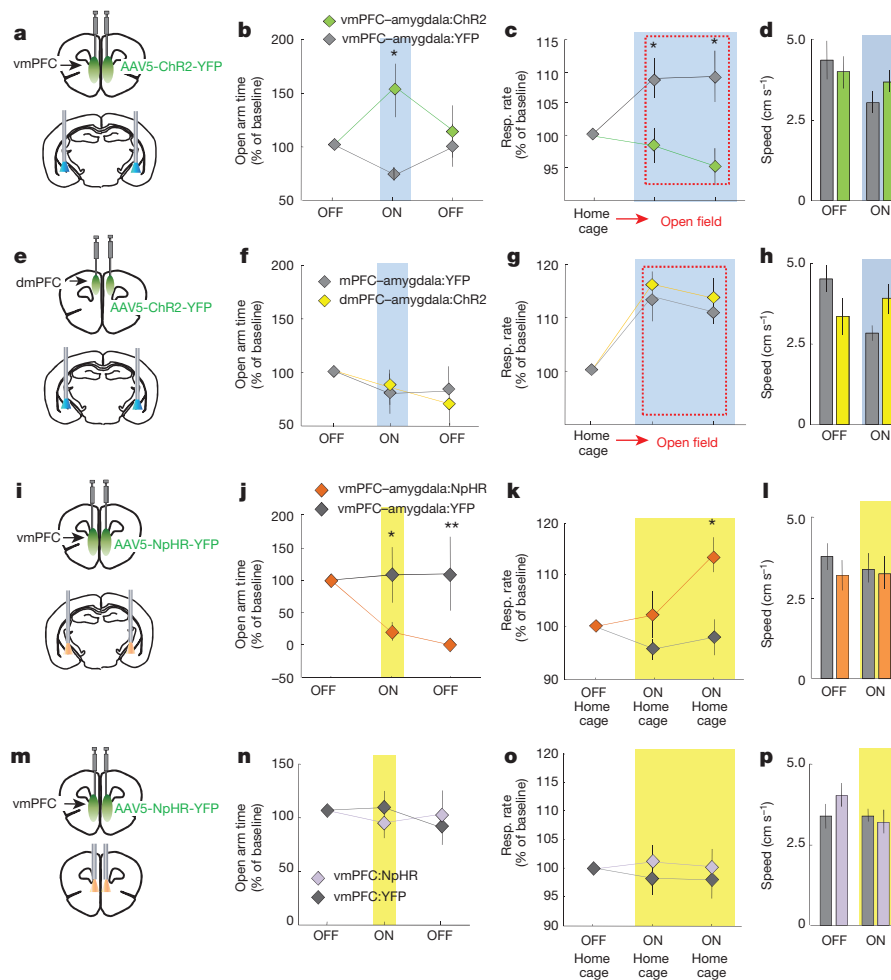
avoidance of open arms of the elevated plus maze (EPM) is a measure of anxiety-related behaviour<sup>24,25</sup>. We expressed channelrhodopsin ChR2-H134R in mouse vmPFC (Extended Data Fig. 1a, b, Supplementary Note 1) and implanted fibre-optic cannulae above the amygdala (vmPFC–amygdala:ChR2 mice, Fig. 1a) or above vmPFC (vmPFC–C:ChR2 mice). vmPFC–amygdala activation decreased avoidance of open spaces in EPM (two-way repeated measures analysis of variance (ANOVA), opsin  $\times$  epoch interaction,  $F_{2,37} = 3.1$ ,  $P = 0.04$ , post-hoc Wilcoxon rank-sum test  $P = 0.009$ ; Fig. 1b, see Extended Data Table 1 for absolute values) and open field (Extended Data Fig. 2d). Projection-targeting appeared to resolve a distinct cell population, since directly exciting the entire vmPFC was not anxiolytic (Extended Data Fig. 2b, c).

Respiratory rates increased during exploration of the anxiogenic open field (relative to home cage)<sup>24,26,27</sup> in control animals (second and third epochs in Fig. 1c, g marked by a dashed red line, Fig. 1c; see Extended Data Fig. 3 for heart rate (a), sham-injected animals (e–h) and raw traces (i, j)). vmPFC–amygdala activation suppressed this increase in respiratory rate (Fig. 1c; two-way repeated measures ANOVA, main effect of opsin,  $F_{2,29} = 11.7$ ,  $P = 0.0014$ , post-hoc Wilcoxon rank-sum test,  $P = 0.03$ ) without changing locomotion (Fig. 1d), or respiratory rate in the home cage (Extended Data Fig. 3b), suggesting suppression of respiratory rate only in anxiogenic settings (Supplementary Note 2). Antidromic spikes in PFC were not readily detected following optogenetic excitation of vmPFC–amygdala:ChR2 terminals (Extended Data Fig. 4). Dorsal mPFC (dmPFC) was investigated next. dmPFC–amygdala:YFP mice did not exhibit labelled fibres in hypothalamus, in contrast to vmPFC–amygdala:YFP mice (Extended Data Fig. 1c, g), supporting correct anatomical targeting<sup>9</sup> (Supplementary Note 1). Upon optical stimulation, dmPFC–amygdala:ChR2 mice showed no changes in behavioural (Fig. 1f, h and Extended Data Fig. 2g) or physiological (Fig. 1g and Extended Data Fig. 2h–j) measures of anxiety; consistent with prior reports<sup>28,29</sup>.

To evaluate the necessity of endogenous vmPFC–amygdala activity for anxiolysis, we expressed the inhibitory halorhodopsin

<sup>1</sup>Department of Bioengineering, Stanford University, Stanford, California 94305, USA. <sup>2</sup>CNC Program, Stanford University, Stanford, California 94304, USA. <sup>3</sup>Neurosciences Program, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA. <sup>5</sup>Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA.

\*These authors contributed equally to this work.



**Figure 1 | Activating vmPFC but not dmPFC terminals in amygdala decreases anxiety.** **a**, vmPFC-amy:ChR2 mice expressing ChR2 in vmPFC with fibre optics above amygdala. **b**, Blue light increased exploration of open arms in vmPFC-amy:ChR2 mice.  $n = 11$  vmPFC-amygdala:ChR2; 10 vmPFC-amygdala:YFP mice. **c**, Respiratory rates recorded for 5 min in home cage and 10 min in open field (red dashed rectangle). Blue light in vmPFC-amygdala:ChR2 mice prevented increases in respiratory rate in open field (OFT) relative to home cage, without altering locomotion (**d**).  $n = 11$  vmPFC-amygdala:ChR2; 6 vmPFC-amygdala:YFP mice. **e–h**, Same as **a–d**, but for dmPFC-amygdala projections.  $n = 7$  dmPFC-amygdala:YFP; 7 mPFC-amygdala:ChR2 mice. **i**, vmPFC-amygdala:NpHR mice expressing eNpHR3.0 (abbreviated in all figures

eNpHR3.0<sup>12,24</sup> in vmPFC (Fig. 1i). vmPFC-amygdala inhibition increased avoidance of open spaces in vmPFC-amygdala:eNpHR3.0 mice in EPM (Fig. 1j; two-way repeated measures ANOVA, main effect of opsin  $\times$  epoch interaction,  $F_{2,25} = 5.0$ ,  $P = 0.014$ , post hoc Wilcoxon rank-sum test  $P = 0.02$ ) and open field (Extended Data Fig. 2e). Moreover, vmPFC-amygdala inhibition in the home cage increased respiratory rate (Fig. 1k), but not locomotion (Fig. 1l) or heart rate (Extended Data Fig. 3d). Conversely, nonspecific mPFC inhibition did not change avoidance of open spaces (Fig. 1m, n and Extended Data Fig. 2f) or respiratory rate (Fig. 1o), revealing that vmPFC-amygdala endogenous activity regulates anxiety states.

### vmPFC-amygdala promotes fear extinction

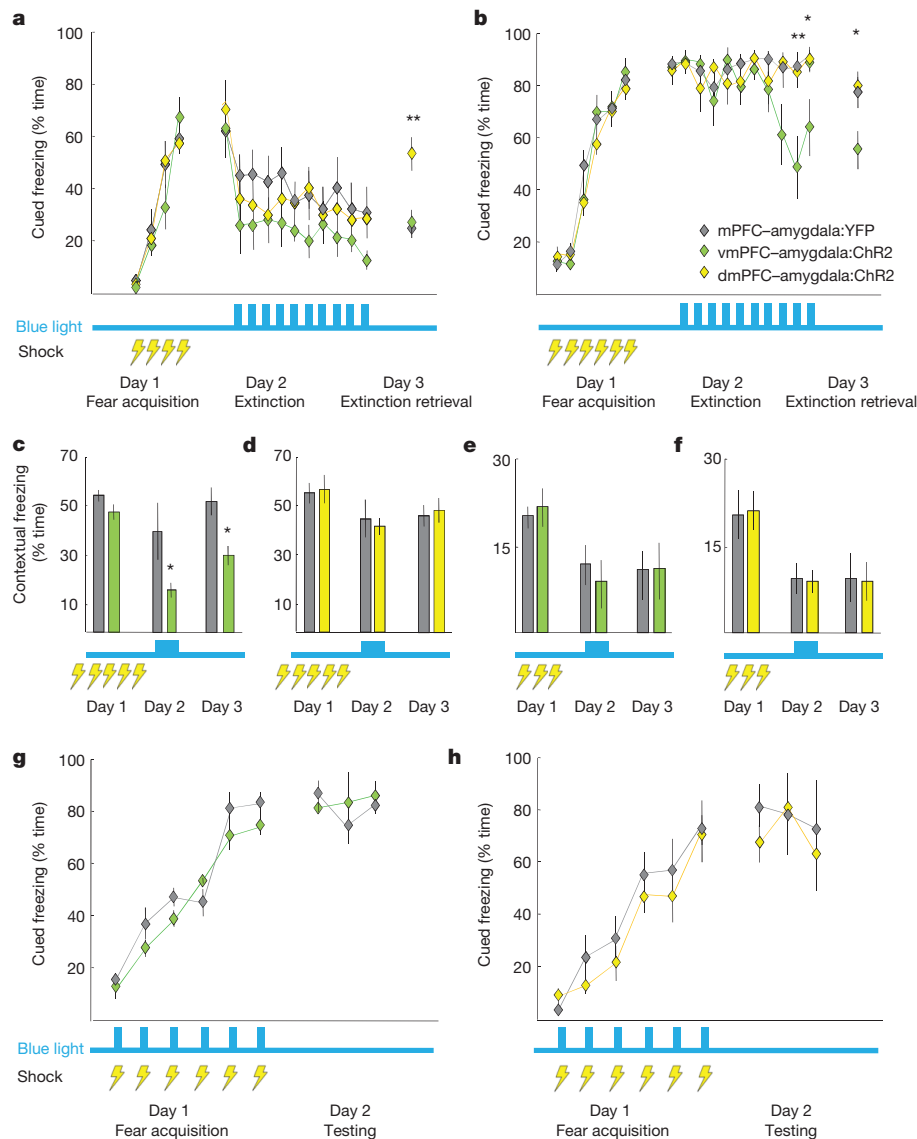
To investigate if this top-down projection controlled fear-conditioning, we employed mild cued fear-conditioning to avoid ceiling effects on freezing. vmPFC-amygdala activation during extinction increased freezing in dmPFC-amygdala:ChR2 mice during extinction retrieval (Fig. 2a,  $P = 0.01$ , Wilcoxon rank-sum test). We next employed stronger cued fear-conditioning to eliminate floor effects, allowing detection

as NpHR) in vmPFC with fibre optics above amygdala. **j, k**, Yellow light decreased exploration of open arms (**j**;  $n = 8$  vmPFC-amygdala:NpHR; 7 vmPFC-amygdala:YFP) and increased home cage respiratory rates (**k**;  $n = 6$  vmPFC-amygdala:NpHR; 6 vmPFC-amygdala:YFP). **l**, Yellow light did not alter overall locomotion ( $n = 14$  vmPFC-amygdala:NpHR; 11 vmPFC-amygdala:YFP). **m**, vmPFC:NpHR mice with fibre optics above vmPFC. **n–p**, Yellow light did not change exploration of EPM (**n**), home cage respiratory rates (**o**) or speed (**p**).  $n = 11$  vmPFC:NpHR; 8 vmPFC:YFP mice. \* $P < 0.05$ ; \*\* $P < 0.01$ , Two-sided Wilcoxon test. Data plotted in 5-min consecutive intervals. Light delivery epochs labelled ON or with blue/yellow bars. Error bars, mean  $\pm$  s.e.m.

of freezing suppression. vmPFC-amygdala activation enhanced cued fear extinction (two-way repeated-measures ANOVA,  $F_{18,250} = 10.73$ ,  $P = 0.0012$  for group type) and extinction retrieval ( $P = 0.04$ , Wilcoxon rank sum test; Fig. 2b); increased extinction retrieval was also seen following stimulation of vmPFC somata (Extended Data Fig. 2a)<sup>7</sup>. vmPFC-amygdala activation also decreased freezing in contextual fear extinction and extinction retrieval (Fig. 2c, two-way repeated-measures ANOVA; significant effect of group type,  $F_{2,29} = 8.7$ ,  $P = 0.009$ ). In contrast, dmPFC-amygdala activation did not alter contextual freezing (Fig. 2d, f).

Behavioural effects (Figs 1 and 2) were not attributable to activation of fibres-of-passage through amygdala, as vmPFC fibres were not observed in posterior sections of amygdala, revealing instead more proximal termination (Extended Data Fig. 5a, b); moreover, photo-stimulation even 500  $\mu$ m posterior to the coordinates used in the above experiments did not modulate anxiety or freezing (Extended Data Fig. 5d, e). These data reveal the vmPFC-amygdala projection suppresses fear-related freezing and high-anxiety states, while the dmPFC-amygdala projection selectively antagonizes only





**Figure 2 | Activation of vmPFC-amygdala and dmPFC-amygdala projections: opposite effects on cued fear.** **a**, Mice fear conditioned to four tone-shock pairings (0.4 mA/1-s shocks) on day 1; extinction on day 2. Blue light delivered on day 2; trials 2–11. dmPFC-amygdala:ChR2 mice froze more than controls during day 3.  $n = 11$  mPFC-amygdala:YFP, 7 vmPFC-amygdala:ChR2, 13 dmPFC-amygdala:ChR2 mice. **b**, Same as **a**, but six tone-shock pairings (0.7 mA/2-s shocks). vmPFC-amygdala:ChR2 mice froze less than controls at end of extinction and during extinction retrieval.  $n = 12$  mPFC-amygdala:YFP, 9 vmPFC-amygdala:ChR2, 10 dmPFC-amygdala:ChR2 mice. **c**, Mice received five foot shocks (0.4 mA/1-s shocks) during contextual fear acquisition. Light decreased

freezing in vmPFC-amygdala:ChR2 mice.  $n = 8$  mPFC-amygdala:YFP; 9 vmPFC-amygdala:ChR2 mice. **d**, Same as **c**, but dmPFC-amygdala mice ( $n = 8$  dmPFC-amygdala:YFP; 8 dmPFC-amygdala:ChR2 mice). **e**, **f**, Contextual fear conditioning: three shocks, 0.4 mA/1 s. Delivery of light during day 2 did not alter freezing in vmPFC-amygdala:ChR2 (**e**) or dmPFC-amygdala:ChR2 (**f**) mice. **e**,  $n = 7$  vmPFC-amygdala:ChR2; 6 vmPFC-amygdala:YFP mice. **f**,  $n = 7$  dmPFC-amygdala:ChR2; 7 dmPFC-amygdala:YFP mice. **g**, **h**, Delivery of light during fear acquisition (0.7 mA/2-s shocks) did not alter freezing in vmPFC-amygdala:ChR2 (**g**) or dmPFC-amygdala:ChR2 (**h**) mice. **g**, **h**,  $n = 8$ , all cohorts;  $*P < 0.05$ ,  $\#P < 0.01$ ; two-sided Wilcoxon test. Error bars, mean  $\pm$  s.e.m.

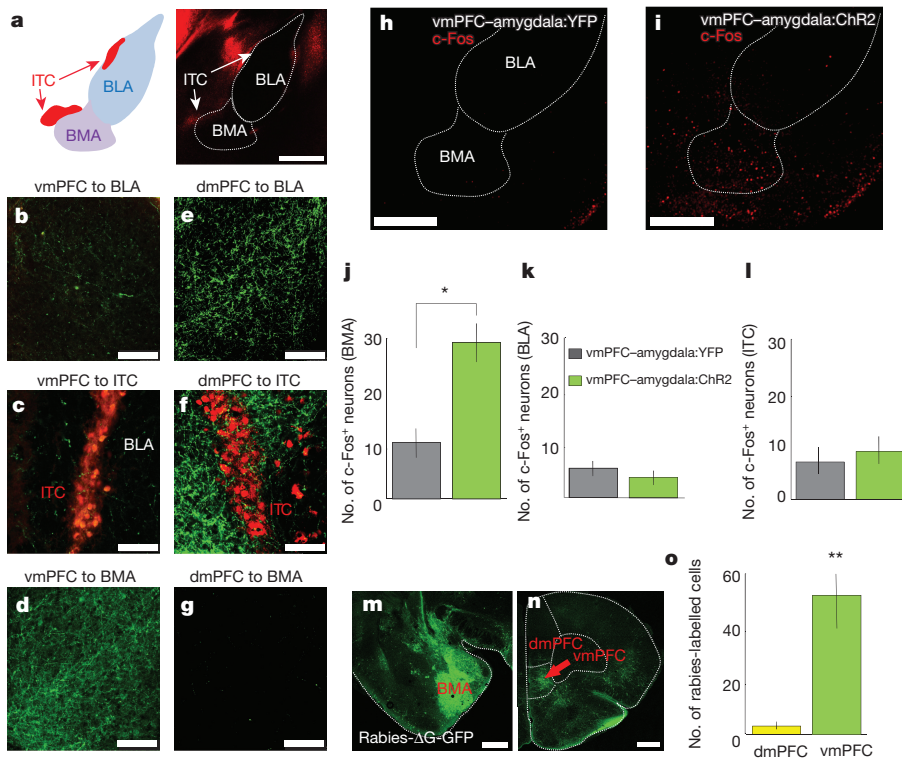
cued fear-extinction. Activation of either dmPFC or vmPFC projections to amygdala during cued fear acquisition did not alter freezing (Fig. 2g, h), in agreement with reports showing lack of mPFC involvement in fear acquisition<sup>30,31</sup>.

### BMA mediates vmPFC top-down control

We next sought precise circuit targets underlying this top-down modulation. We observed sparse vmPFC fibres in BLA and ITCs (Fig. 3b, c; CLARITY data, Supplementary Video 1; Extended Data Fig. 5c), but prominent investment of basomedial amygdala (BMA; Fig. 3d)<sup>9,32,33</sup>. Conversely, dmPFC fibres strongly invested BLA (and ITCs, not previously reported), while sparing the BMA (Fig. 3e–g). Accordingly, vmPFC-amygdala activation induced c-Fos in BMA, but not in BLA or ITCs (Fig. 3h–l). We next injected the retrogradely propagating virus

rabies-ΔG-GFP into BMA and identified a strong input specifically from vmPFC (Fig. 3m–o), predominantly from layer V (Extended Data Fig. 6a–c); similar results were obtained with retrobeads (Extended Data Fig. 6d–h). Finally, we obtained whole-cell recordings from amygdala cells during optical stimulation of ChR2-expressing mPFC terminals (in tetrodotoxin and 4-aminopyridine to abolish polysynaptic responses); stimulation of vmPFC terminals elicited monosynaptic responses in BMA (7/11 cells), in contrast to BLA (1/8 cells) or ITCs (0/19 cells) (Fig. 4a–c). Conversely, dmPFC terminals spared BMA but elicited monosynaptic responses in BLA and the ventral ITC cluster (Fig. 4d–f).

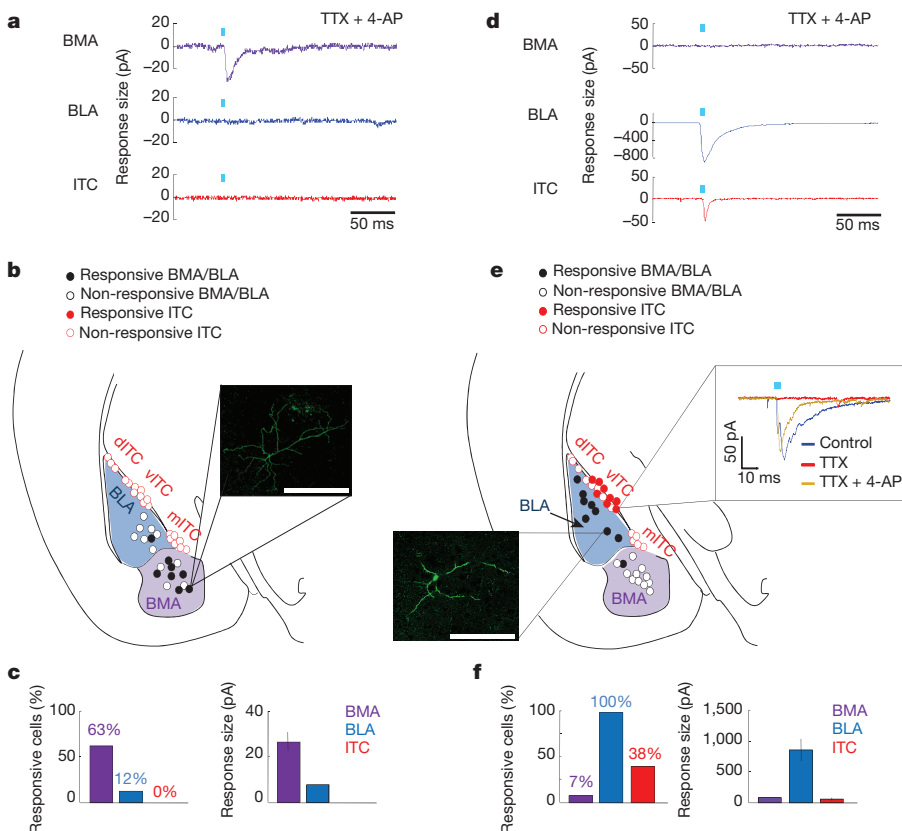
We next found that BMA cells *in vitro* could follow 10 Hz stimulation of vmPFC fibres (Extended Data Fig. 7a), and that vmPFC-amygdala activation elicited initial BMA excitation followed by inhibition *in vitro*



**Figure 3 | Basomedial amygdala: major target of vmPFC in amygdala.** **a**, Left: location of amygdala nuclei. Right: *lox*-tdTomato reporter line crossed with EY266 D1R::Cre line for ITC visualization. Scale bar, 500  $\mu$ m. **b–d**, AAV5-CamK2 $\alpha$ ::ChR2-YFP injected in vmPFC; vmPFC fibres imaged in BLA (**b**), ITC (**c**) and BMA (**d**). **a–d**, Representative images from one mouse chosen out of  $n = 7$  mice. **e–g**, Same as **b–d**, but for dmPFC. **e–g**, Representative images from one mouse chosen out of  $n = 7$  mice. Scale bars, 50  $\mu$ m (**c**, **f**); 100  $\mu$ m (**b**, **e**, **d**, **g**). **h–l**, c-Fos-expressing cells (red) counted in BMA (**j**), BLA (**k**) and ITCs (**l**) following vmPFC-amygdala activation; increased c-Fos can be seen in BMA of vmPFC-amygdala:ChR2 mice.  $n = 5$  mice, for each group, 4 slices per mouse. **m–o**, Rabies- $\Delta$ G-GFP injected in BMA (**m**); retrogradely labelled cells counted (**n–o**).  $n = 4$  mice (**o**). Scale bar, 500  $\mu$ m (**h**, **i**, **m**, **n**). \* $P < 0.05$ , \*\* $P < 0.01$ , two-sided Wilcoxon test. Error bars, mean  $\pm$  s.e.m.

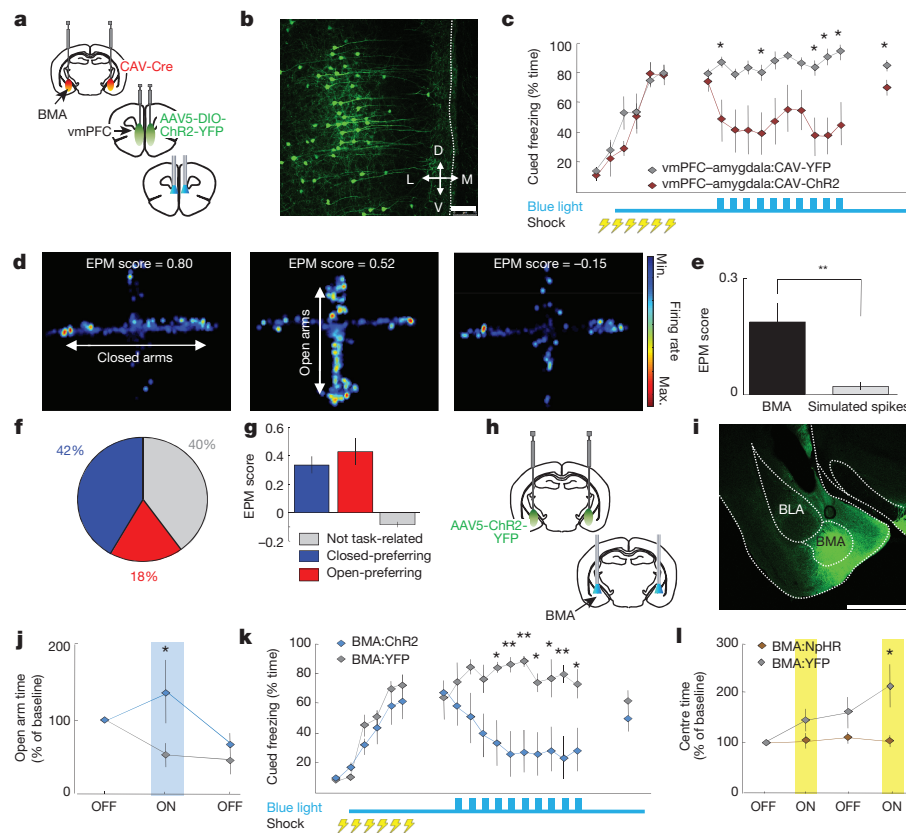
(Extended Data Fig. 7b) and *in vivo* (Extended Data Fig. 7c). The inhibition was probably mediated by directly recruited GAD2-expressing cells in BMA, since 91% of these receive vmPFC inputs (Extended Data Fig. 7g). Finally, we tested if specific activation of vmPFC cell bodies that project to BMA suppressed freezing. We targeted BMA-projecting

vmPFC cells by injecting the retrogradely propagating canine adenovirus (CAV) encoding Cre-recombinase into BMA (Extended Data Fig. 8, Supplementary Note 3) and directing AAV-DIO-ChR2-YFP as well as optical fibres into mPFC (Fig. 5a, b). Activation of these cells decreased cued freezing during extinction and extinction retrieval



**Figure 4 | Functional connectivity: mPFC inputs to amygdala.** **a**, Example traces from one mouse showing stimulation of ChR2-expressing vmPFC terminals in amygdala (acute slice) elicited responses in BMA, but not BLA or ITCs. Recordings in TTX and 4-AP to abolish polysynaptic responses. **b**, Locations of recorded cells. Inset: responsive biocytin-filled BMA cell. dITC, vITC and mITC: dorsal, ventral and main ITC clusters.  $n = 11$  BMA, 8 BLA and 19 ITC cells. **c**, Mean percentage of responsive cells and response sizes.  $n = 4$  mice (**a–c**). **d–f**, Same as **a–c**, but dmPFC terminal stimulation.  $n = 4$  mice (**d–f**). **b**, **e**, Inset scale bar, 250  $\mu$ m. Error bars, mean  $\pm$  s.e.m.





**Figure 5 | BMA cells encode anxiety-related contextual features, decrease anxiety and decrease freezing.** **a**, CAV-Cre in BMA and AAV5-DIO-ChR2 in vmPFC. **b**, YFP-expression in BMA-projecting vmPFC neurons. Scale bar, 75  $\mu$ m. **c**, Light-decreased freezing. **b, c**,  $n = 7$  vmPFC-amygdala:CAV-YFP, 8 vmPFC-amygdala:CAV-ChR2. **d**, Heat maps of representative cells preferentially firing in closed arms (left) or open arms (middle), or with no arm-type preference (right). **e**, EPM scores (strong encoding of arm-type) for BMA units versus simulated spike trains;  $n = 3$  single-unit examples chosen from 38 BMA units recorded from  $n = 4$  mice.

**f**, Distribution of open- and closed-arm-preferring cells (EPM score  $> 0$ ) versus no task-related firing (EPM score  $\leq 0$ ). **g**, Average EPM scores. **d–g**,  $n = 38$  BMA single units from  $n = 4$  mice. **h, i**, BMA:ChR2 mice expressing ChR2 in BMA with fibre optics above BMA. Scale bar, 1 mm. **j, k**, BMA activation increased open-arm exploration (**j**) and decreased freezing (**k**). **i, j**,  $n = 8$  BMA:ChR2; 9 BMA:YFP. **k**,  $n = 6$  BMA:ChR2; 7 BMA:YFP. **l**, Delivery of light increased OFT centre-avoidance in BMA:NpHR mice.  $n = 8$  BMA:NpHR; 7 BMA:YFP. \* $P < 0.05$ , \*\* $P < 0.01$ , two-sided Wilcoxon test. Error bars, mean  $\pm$  s.e.m.

(Fig. 5c), demonstrating that BMA can account for the behavioural effects seen following vmPFC-amygdala activation.

### BMA inhibits fear and anxiety

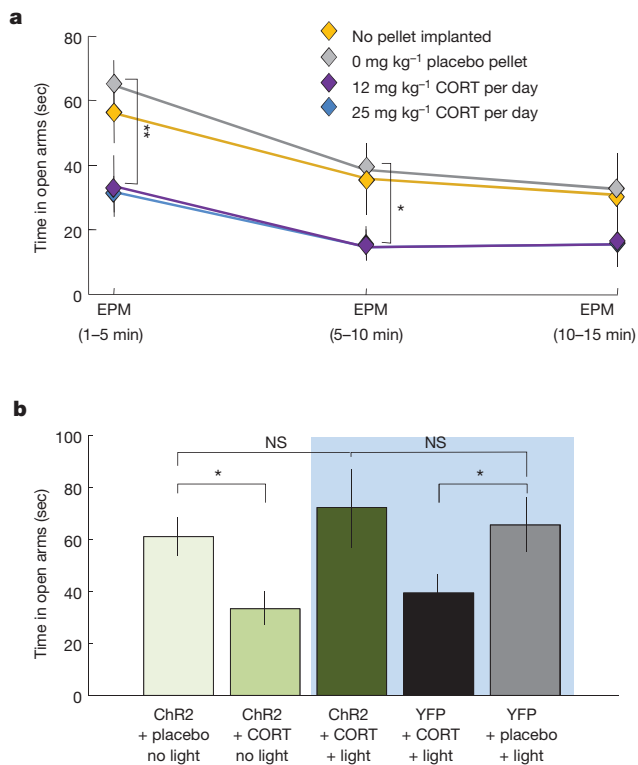
If the BMA is naturally used to suppress fear and anxiety, then BMA spiking should differentiate safe and aversive contexts. We recorded BMA cells (Extended Data Fig. 9a, b) and used EPM scores to quantify arm-type encoding (Methods and refs 24,34). BMA neurons exhibited stable firing patterns (Extended Data Fig. 9c) with higher EPM scores than would be expected by chance (Fig. 5d, e;  $P = 0.01$ , Wilcoxon rank-sum test). In light–dark tests (LDT), where anxiety is associated with the light compartment of the apparatus<sup>35</sup>, the same BMA cells showed similar changes in firing rate in EPM open arms and LDT light compartments (Extended Data Fig. 9d), indicating that BMA neurons encode anxiety-associated contextual features. Among cells with positive EPM scores, neurons that fired preferentially in closed arms were over-represented ( $P < 0.05$ , binomial test) relative to open-arm-preferring cells (Fig. 5f). Similarly, a larger fraction of BMA cells ( $P < 0.01$ , binomial test) displayed inhibition versus excitation to fear-conditioned tones (Extended Data Fig. 9e). Consistent with this abundance of low-spiking BMA cells in actively aversive/anxiogenic settings, and the hypothesized countervailing role for excitatory anxiolytic inputs to BMA, we found BMA activation (Fig. 5h, i) decreased avoidance of open arms (Fig. 5j; two-way repeated-measures ANOVA, main effect of opsin,  $F_{2,29} = 4.9$ ,  $P = 0.03$ , post hoc Wilcoxon rank-sum test  $P = 0.04$ ) and freezing in cued fear (Fig. 5k; two-way repeated-measures ANOVA, main

effect of opsin,  $F_{9,99} = 12.3$ ,  $P = 0.004$ ). Conversely, inhibition of BMA with eNpHR3.0 was anxiogenic in the open field ( $P < 0.05$ , main effect of opsin, two-way repeated-measures ANOVA,  $F_{2,25} = 4.1$ ,  $P < 0.01$ , Fig. 5l), perhaps concordant with a role for the BMA in fear<sup>36,37</sup>.

Anxiolysis by the vmPFC–BMA circuit raised the possibility of top-down modulation of behaviour elicited by pathological conditions (such as high levels of corticosterone, since elevated cortisol-related stress hormones are known to be associated with anxiety disorders<sup>38–40</sup>). Indeed, chronic corticosterone treatment increased<sup>41,42</sup> avoidance of open arms (Fig. 6a,  $P = 0.018$ , Wilcoxon rank-sum test), an effect reversed by vmPFC-amygdala activation (Fig. 6b).

### Discussion

Here we identify the BMA as the direct target of vmPFC that suppresses fear-related freezing and anxiety states. Although these and other regions have been studied separately, a precise direct target for well-resolved vmPFC projections in top-down control of anxiety and fear had not been identified, nor had roles for the vmPFC–BMA projection been demonstrated. We find that this pathway is sufficient and necessary for anxiolysis; that cells inhibited by aversive stimuli are over-represented in BMA, and that BMA activity suppresses freezing and elevated-anxiety states. We also show the major source of prefrontal input to ITCs is not vmPFC, but rather dmPFC; for such circuit-level insight, precise projection-validated targeting of vmPFC versus dmPFC was crucial (Supplementary Note 1).



**Figure 6 | vmPFC-BMA projection reverses anxiogenic effect of corticosterone.** **a**, Chronic corticosterone (CORT) treatment decreased open-arm exploration.  $n = 10$  no-pellet, 10 placebo pellet, 7 (12 mg kg<sup>-1</sup>) CORT, 8 (25 mg kg<sup>-1</sup>) CORT mice. **b**, vmPFC-amygdala activation reversed effect of CORT. NS, not significant.  $n = 15$  YFP placebo + light, 15 YFP CORT + light, 15 ChR2 CORT + light, 7 ChR2 CORT + no light, 8 ChR2 placebo + no light; \* $P < 0.05$ , \*\* $P < 0.01$ , two-sided Wilcoxon test. Error bars, mean ± s.e.m.

While vmPFC-BMA activation suppressed freezing and elevated-anxiety states, neither bulk vmPFC activation (Extended Data Fig. 2a–c), nor control of other cortical (dmPFC) afferents to the target (amygdala; Fig. 1e–h), gave rise to similar effects. The lack of anxiolysis following direct vmPFC activation points to diversity in the vmPFC population, which probably includes cells that project to regions counteracting anxiolytic effects of vmPFC-BMA projections (Supplementary Note 4). Interestingly, inhibition of the vmPFC-amygdala projection increased respiration but not heart rate, as seen in the bed nucleus of the stria terminalis (BNST)<sup>24</sup>, suggesting that heart rate may be regulated by other mechanisms perhaps related to direct sympathetic activation.

vmPFC also innervates the cortical and medial amygdala<sup>9</sup>; these structures may be activated in vmPFC-amygdala:ChR2 mice, and are not ruled out in regulating anxiolysis. Nevertheless, direct activation of BMA inhibits freezing. Interestingly, the BMA projects to several areas implicated in anxiolysis, such as the central lateral amygdala<sup>43</sup>, the vmPFC, and the BNST anterodorsal nucleus<sup>43</sup>, but not to directly adjacent anxiogenic regions, such as the dmPFC and the BNST oval nucleus (Extended Data Fig. 9k,l)<sup>24,43</sup>. Although freezing was not increased by direct optical inhibition of BMA cell bodies (Extended Data Fig. 9j), pharmacological inhibition of the BMA increases freezing<sup>37</sup>, and its activation inhibits social stress-induced cardiovascular responses<sup>44</sup>.

The minimal vmPFC-ITC connectivity observed in mice (Figs 3c and 4c) does not conflict with studies in rats showing increased ITC activity following electrical<sup>45</sup> or pharmacological vmPFC activation<sup>46</sup>, in which no direct projections were tested (such effects could be implemented through indirect routes such as via dmPFC, which we find directly projects to ITCs (Figs 3f and 4f). Influence of vmPFC

on ITCs is not ruled out (despite the low functional<sup>47</sup> and anatomical<sup>8</sup> connectivity between these regions), and it remains possible that vmPFC-BMA and vmPFC-ITC influences could play complementary roles in regulating distinct fear or anxiety phenotypes. Diverse local direct target possibilities for mediating top-down control of fear have been hypothesized, including ITCs or BLA<sup>48</sup>; here we directly identify top-down control anatomy, as activation of BMA-projecting vmPFC neurons inhibits freezing (Fig. 5c).

A final intriguing aspect is that while direct BMA activation increased extinction, the extinction memory did not persist (Fig. 5k), whereas driving the vmPFC-BMA projection gave rise to stable extinction (Fig. 2b). Moreover, vmPFC-BMA stimulation only reduced freezing with several fear-conditioned tone presentations (Fig. 2b), suggesting that pairing cue exposure with vmPFC-BMA activity induces lasting plasticity. Such plasticity in vmPFC projections for lasting effects on fear memory is potentially relevant to therapeutic strategies and to natural anxiety and fear regulation.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 December 2014; accepted 4 September 2015.

Published online 4 November 2015.

1. Tromp, D. P. et al. Reduced structural connectivity of a major frontolimbic pathway in generalized anxiety disorder. *Arch. Gen. Psychiatry* **69**, 925–934 (2012).
2. Prater, K. E., Hosanagar, A., Klumpp, H., Angstadt, M. & Phan, K. L. Aberrant amygdala-frontal cortex connectivity during perception of fearful faces and at rest in generalized social anxiety disorder. *Depress. Anxiety* **30**, 234–241 (2013).
3. Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S. L. & Quirk, G. J. Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learn. Mem.* **13**, 728–733 (2006).
4. Knapska, E. & Maren, S. Reciprocal patterns of c-Fos expression in the medial prefrontal cortex and amygdala after extinction and renewal of conditioned fear. *Learn. Mem.* **16**, 486–493 (2009).
5. LeDoux, J. The amygdala. *Curr. Biol.* **17**, R868–R874 (2007).
6. Ji, G. & Neugebauer, V. Modulation of medial prefrontal cortical activity using *in vivo* recordings and optogenetics. *Mol. Brain* **5**, 36 (2012).
7. Do-Monte, F. H., Manzano-Nieves, G., Quinones-Laracuente, K., Ramos-Medina, L. & Quirk, G. J. Revisiting the role of infralimbic cortex in fear extinction with optogenetics. *J. Neurosci.* **35**, 3607–3615 (2015).
8. Pinard, C. R., Mascagni, F. & McDonald, A. J. Medial prefrontal cortical innervation of the intercalated nuclear region of the amygdala. *Neuroscience* **205**, 112–124 (2012).
9. Vertes, R. P. Differential projections of the infralimbic and prelimbic cortex in the rat. *Synapse* **51**, 32–58 (2004).
10. Sierra-Mercado, D., Padilla-Coreano, N. & Quirk, G. J. Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology* **36**, 529–538 (2011).
11. Maren, S., Aharonov, G., Stote, D. L. & Fanselow, M. S. N-methyl-D-aspartate receptors in the basolateral amygdala are required for both acquisition and expression of conditional fear in rats. *Behav. Neurosci.* **110**, 1365–1374 (1996).
12. Tye, K. M. et al. Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature* **471**, 358–362 (2011).
13. Likhtik, E., Popa, D., Apergis-Schoute, J., Fidacaro, G. A. & Pare, D. Amygdala intercalated neurons are required for expression of fear extinction. *Nature* **454**, 642–645 (2008).
14. Likhtik, E., Pelletier, J. G., Paz, R. & Pare, D. Prefrontal control of the amygdala. *J. Neurosci.* **25**, 7429–7437 (2005).
15. Rosenkranz, J. A. & Grace, A. A. Cellular mechanisms of infralimbic and prelimbic prefrontal cortical inhibition and dopaminergic modulation of basolateral amygdala neurons *in vivo*. *J. Neurosci.* **22**, 324–337 (2002).
16. Hubner, C., Bosch, D., Gall, A., Luthi, A. & Ehrlich, I. *Ex vivo* dissection of optogenetically activated mPFC and hippocampal inputs to neurons in the basolateral amygdala: implications for fear and emotional memory. *Front. Behav. Neurosci.* **8**, 64 (2014).
17. Cho, J. H., Deisseroth, K. & Bolshakov, V. Y. Synaptic encoding of fear extinction in mPFC-amygdala circuits. *Neuron* **80**, 1491–1507 (2013).
18. Jinks, A. L. & McGregor, I. S. Modulation of anxiety-related behaviours following lesions of the prelimbic or infralimbic cortex in the rat. *Brain Res.* **772**, 181–190 (1997).
19. Sullivan, R. M. & Gratton, A. Behavioral effects of excitotoxic lesions of ventral medial prefrontal cortex in the rat are hemisphere-dependent. *Brain Res.* **927**, 69–79 (2002).



20. Bi, L. L. *et al.* Enhanced excitability in the infralimbic cortex produces anxiety-like behaviors. *Neuropharmacology* **72**, 148–156 (2013).
21. Wall, P. M. & Messier, C. U-69,593 microinjection in the infralimbic cortex reduces anxiety and enhances spontaneous alternation memory in mice. *Brain Res.* **856**, 259–280 (2000).
22. Stevenson, C. W. Role of amygdala-prefrontal cortex circuitry in regulating the expression of contextual fear memory. *Neurobiol. Learn. Mem.* **96**, 315–323 (2011).
23. Maaswinkel, H., Gispen, W. H. & Spruijt, B. M. Effects of an electrolytic lesion of the prelimbic area on anxiety-related and cognitive tasks in the rat. *Behav. Brain Res.* **79**, 51–59 (1996).
24. Kim, S. Y. *et al.* Diverging neural pathways assemble a behavioural state from separable features in anxiety. *Nature* **496**, 219–223 (2013).
25. File, S. E., Gonzalez, L. E. & Andrews, N. Comparative study of pre- and postsynaptic 5-HT<sub>1A</sub> receptor modulation of anxiety in two ethological animal tests. *J. Neurosci.* **16**, 4810–4815 (1996).
26. Martin, B. The assessment of anxiety by physiological behavioral measures. *Psychol. Bull.* **58**, 234–255 (1961).
27. Suess, W. M., Alexander, A. B., Smith, D. D., Sweeney, H. W. & Marion, R. J. The effects of psychological stress on respiration: a preliminary study of anxiety and hyperventilation. *Psychophysiology* **17**, 535–540 (1980).
28. Corcoran, K. A. & Quirk, G. J. Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears. *J. Neurosci.* **27**, 840–844 (2007).
29. Rebello, T. J. *et al.* Postnatal day 2 to 11 constitutes a 5-HT-sensitive period impacting adult mPFC function. *J. Neurosci.* **34**, 12379–12393 (2014).
30. Sierra-Mercado, D. Jr, Corcoran, K. A., Lebron-Milad, K. & Quirk, G. J. Inactivation of the ventromedial prefrontal cortex reduces expression of conditioned fear and impairs subsequent recall of extinction. *Eur. J. Neurosci.* **24**, 1751–1758 (2006).
31. Quirk, G. J., Russo, G. K., Barron, J. L. & Lebron, K. The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *J. Neurosci.* **20**, 6225–6231 (2000).
32. Cassell, M. D. & Wright, D. J. Topography of projections from the medial prefrontal cortex to the amygdala in the rat. *Brain Res. Bull.* **17**, 321–333 (1986).
33. Hurley, K. M., Herbert, H., Moga, M. M. & Saper, C. B. Efferent projections of the infralimbic cortex of the rat. *J. Comp. Neurol.* **308**, 249–276 (1991).
34. Adhikari, A., Topiwala, M. A. & Gordon, J. A. Single units in the medial prefrontal cortex with anxiety-related firing patterns are preferentially influenced by ventral hippocampal activity. *Neuron* **71**, 898–910 (2011).
35. Misslin, R., Belzung, C. & Vogel, E. Behavioural validation of a light/dark choice procedure for testing anti-anxiety agents. *Behav. Processes* **18**, 119–132 (1989).
36. Martinez, R. C., Carvalho-Netto, E. F., Ribeiro-Barbosa, E. R., Baldo, M. V. & Canteras, N. S. Amygdalar roles during exposure to a live predator and to a predator-associated context. *Neuroscience* **172**, 314–328 (2011).
37. Amano, T., Duvarci, S., Popa, D. & Pare, D. The fear circuit revisited: contributions of the basal amygdala nuclei to conditioned fear. *J. Neurosci.* **31**, 15481–15489 (2011).
38. Chaudieu, I. *et al.* Abnormal reactions to environmental stress in elderly persons with anxiety disorders: evidence from a population study of diurnal cortisol changes. *J. Affect. Disord.* **106**, 307–313 (2008).
39. Dimopoulou, C. *et al.* Increased prevalence of anxiety-associated personality traits in patients with Cushing's disease: a cross-sectional study. *Neuroendocrinology* **97**, 139–145 (2013).
40. Vreeburg, S. A. *et al.* Salivary cortisol levels in persons with and without different anxiety disorders. *Psychosom. Med.* **72**, 340–347 (2010).
41. Gregus, A., Wintink, A. J., Davis, A. C. & Kalynchuk, L. E. Effect of repeated corticosterone injections and restraint stress on anxiety and depression-like behavior in male rats. *Behav. Brain Res.* **156**, 105–114 (2005).
42. Lee, B. *et al.* Chronic administration of catechin decreases depression and anxiety-like behaviors in a rat model using chronic corticosterone injections. *Biomol. Ther. (Seoul)* **21**, 313–322 (2013).
43. Petrovich, G. D., Risold, P. Y. & Swanson, L. W. Organization of projections from the basomedial nucleus of the amygdala: a PHAL study in the rat. *J. Comp. Neurol.* **374**, 387–420 (1996).
44. De Abreu, A. R. *et al.* Activation of the basomedial amygdala suppresses the cardiovascular response to an emotional stress in rats. In *Society for Neuroscience Annual Meeting 2014* (Society for Neuroscience, 2014).
45. Amir, A., Amano, T. & Pare, D. Physiological identification and infralimbic responsiveness of rat intercalated amygdala neurons. *J. Neurophysiol.* **105**, 3054–3066 (2011).
46. Berretta, S. & Pantazopoulos, H., Caldera, M., Pantazopoulos, P. & Pare, D. Infralimbic cortex activation increases c-Fos expression in intercalated neurons of the amygdala. *Neuroscience* **132**, 943–953 (2005).
47. Strobel, C., Marek, R., Gooch, H.M., Sullivan, R.K. & Sah, P. Prefrontal and auditory input to intercalated neurons of the amygdala. *Cell Rep.* **10**, 1435–1442 (2015).
48. Bukalo, O. *et al.* Prefrontal inputs to the amygdala instruct fear extinction memory formation. *Sci. Adv.* **1**, e1500251 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank B. K. Lim and the Malenka laboratory for providing Rabies-ΔG-GFP. We thank the Deisseroth laboratory for helpful discussions. A.A. is supported by the Walter V. and Idun Berry award, a K99 award (NIMH K99MH106649) and a NARSAD Young Investigator fellowship. T.L. is a NRSA (1F32MH105053-01) fellow. K.D. is supported by NIMH, the DARPA Neuro-FAST program, NIDA, Peter and Ann Tarlton, NSF, the Simons Foundation, the Gatsby Foundation, the Wieggers Family Fund, the Nancy and James Grosfeld Foundation, the H. L. Snyder Medical Foundation, the Samuel and Betsy Reeves Fund, the Vincent V. C. Woo Fund, and the Albert Yu and Mary Bechman Foundations. All optogenetics (<http://www.optogenetics.org>) and CLARITY (<http://clarityresourcecenter.org>) reagents and protocols are distributed and supported freely.

**Author Contributions** A.A., T.N.L., J.F., L.A.G., S.P., E.F., J.J.M., L.Y., T.J.D., J.H.J., S.-Y.K. and A.L. collected the data. A.A., T.N.L. and J.F. contributed equally. A.A., T.N.L., J.F. and K.D. designed experiments. A.A., T.N.L. and J.F. analysed the data. K.D. supervised all aspects of the work. A.A., T.N.L., J.F. and K.D. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.D. ([deissero@stanford.edu](mailto:deissero@stanford.edu)).

## METHODS

**Subjects: mice.** Male and female C57BL/6 mice were group-housed in a reverse 12 h light/dark cycle. Mice were 8 to 12 weeks old at the time of viral infusion. Food and water were given *ad libitum*. Dopamine receptor D1a (Drd1a)-Cre transgenic mice (EY266) were obtained from GENSAT. A19 *lox*-tdTomato mice were purchased from JAX (line 007902). All experimental protocols were approved by the Stanford University Institutional Animal Care and Use Committee and were in accordance with the guidelines from the National Institutes of Health. Sample sizes were chosen based on previous behavioural optogenetics studies on anxiety, which typically use 6–10 mice per group<sup>24</sup>. No statistical methods were used to predetermine sample size. Behavioural assays were conducted with male mice only, while both male and female mice were used for anatomical tracing and *in vitro* electrophysiology assays.

**Subjects: rats.** All data shown are from mice, except for Extended Data Fig. 5c, which shows data from rats. Wild-type male Sprague Dawley rats (250–400 g) were obtained from Charles River and were individually housed on a standard 12 h light/dark cycle and given food and water *ad libitum*.

**Viruses.** All the adeno-associated virus (AAV) vectors used were serotyped with AAV5 coat proteins and packaged by the University of North Carolina Vector Core. All virus stock solutions were diluted to  $2 \times 10^{12}$  particles per ml. For Chr2 and YFP mice, AAV5:CamK2a:Chr2(H134R)-eYFP and AAV5:CamK2a:eYFP were infused, respectively. The CamK2a promoter was used to induce opsin expression preferentially in excitatory projection cells. AAV5-DIO-Efla-mCherry (used in the BMA of GAD2cre mice) and AAV5-DIO-Efla-ChR2(H134R)-YFP (used in vmPFC-amy:CAV-ChR2 mice) were injected at  $8 \times 10^{12}$  particles per ml. The maps for these viral vectors are available at <http://www.optogenetics.org>. Rabies-ΔG-GFP was provided by B. K. Lim and R. C. Malenka. CAV-Cre virus was obtained from The Institute of Molecular Genetics of Montpellier vector core (France), at  $6.2 \times 10^{12}$  particles per ml.

**Surgery: mice.** Eight-week-old mice were anaesthetized with 1.5–3.0% isoflurane and placed in a stereotaxic apparatus (Kopf Instruments). A scalpel was used to open an incision along the midline to expose the skull. After performing a craniotomy, 1.0 μl of AAV5-CamK2a-ChR2(H134R)-YFP at a titre of  $2 \times 10^{12}$  particles per ml was injected per site (vmPFC or dmPFC) using a 10 μl nanofill syringe (World Precision Instruments) at  $0.1 \mu\text{l min}^{-1}$ . The syringe was coupled to a 33 gauge bevelled needle, and the bevel was placed to face the anterior side of the animal. The syringe was slowly retracted 20 min after the start of the infusion. A slow infusion rate followed by 10 min of waiting before retracting the syringe is crucial to restrict viral expression to the vmPFC or dmPFC. Mice in vmPFC and dmPFC groups received unilateral viral infusion and fibre optic cannula implantation (0.22 NA, 200 μm diameter; Doric Lenses) in either the vmPFC or the dmPFC. Infusion coordinates were: anteroposterior, 1.7 mm; mediolateral, 0.25 mm; dorsoventral, 3.3 mm for the vmPFC (injection centered in the dorsal peduncular cortex) and anteroposterior, +1.9 mm; mediolateral, 0.25 mm; dorsoventral, 2.0 mm for the dmPFC. Fibre optic cannula implantation coordinates were: anteroposterior, 1.7 mm; mediolateral, 0.25 mm; dorsoventral, 2.8 mm for the vmPFC and anteroposterior, +1.9 mm; mediolateral, 0.25 mm; dorsoventral, 1.5 mm for the dmPFC. For vmPFC-amygdala and dmPFC-amygdala mice, mPFC virus infusions were done bilaterally and fibreoptic cannulae were implanted bilaterally above amygdala (anteroposterior, –1.3 mm; mediolateral, 3.3 mm; dorsoventral, 4.7 mm). All coordinates were measured from bregma.

Viral infusions targeting the vmPFC were centred in the dorsal peduncular cortex (Supplementary Note 1). These injections produced strong opsin expression in the peduncular cortex and the adjacent infralimbic cortex. On average,  $61 \pm 14\%$  of vmPFC cells expressed YFP (data not shown). Viral infusions targeting the dmPFC were centred in the cingulate cortex. These injections produced strong opsin expression in the cingulate cortex and the adjacent prelimbic cortex. Only mice with opsin expression restricted to either the vmPFC or the dmPFC were used for behavioural assays. Mice with proper injections did not have a substantial number of YFP-expressing cell bodies outside the target region. However, fibres may be visible outside the target region, as the vmPFC and the dmPFC are reciprocally connected<sup>9</sup>. The boundary between dmPFC and vmPFC was determined by comparing which coronal section in a reference brain atlas most closely corresponds to section being imaged. The boundaries were defined based on the position of easily identified structural landmarks, such as the forceps minor of the corpus callosum. Additionally, high magnification confocal imaging ( $20\times$ ) was used to evaluate if YFP-expressing cell bodies was contained in a specific mPFC subregion. The projection patterns from dmPFC and vmPFC-injected mice also differed. Specifically, fibres from the vmPFC were seen in the dorsomedial and posterior nuclei of the hypothalamus, but not in the ventromedial hypothalamus. Fibres from the dmPFC were not found to innervate any part of the hypothalamus noticeably. This pattern of hypothalamic

projections is in good agreement with prior anatomy studies comparing vmPFC and dmPFC projections<sup>9</sup>. Thus, mice injected in the vmPFC showed prominent fibres in the BMA and hypothalamus, but not in the BLA, while mice injected in the dmPFC had the complementary projection pattern. All mice included in the study had both YFP-expressing cell bodies confined to an mPFC subregion and the expected fibre projection pattern.

For mice expressing Chr2-YFP, NpHR-YFP or YFP in the BMA, 0.5 μl of AAV5-CaMK2a:Chr2(H134R)-YFP or AAV5-CaMK2a:eNpHR3.0-YFP at a titre of  $2 \times 10^{12}$  particles per ml was injected in the BMA (anteroposterior, –1.3 mm; mediolateral, 2.8 mm; dorsoventral, 5.6 mm from bregma), and fibre optic cannulae were implanted 0.5 mm above the infusion site. Adhesive cement (C&B metabond) and dental cement (Stoelting) were used to securely attach the fibre optic cannulae to the skull. For experiments employing CAV-Cre, this virus was injected at  $6.2 \times 10^{12}$  particles per ml in the BMA (0.5 μl bilaterally). We also injected 1.0 μl bilaterally of AAV5-DIO-Efla-ChR2-YFP ( $8 \times 10^{12}$  particles per ml titre) in the vmPFC in these mice. For patch-clamping experiments using GAD2-Cre mice, we injected 1.0 μl of AAV5-Efla-mCherry ( $8 \times 10^{12}$  particles per ml titre) in the BMA to label GAD2-expressing BMA interneurons. In these mice we also injected 1.0 μl bilaterally of AAV5-CamK2a-ChR2-YFP ( $4 \times 10^{12}$  particles per ml titre) in the vmPFC. Mice recovered from surgeries in a heated cage before returning to their home cage. Half the mice in each cage were randomly assigned to YFP or Chr2/NpHR groups.

**Surgery: rats.** Stereotaxic surgeries were with rats (8–10 weeks of age) under isoflurane anaesthesia (4% initially, maintained at 2–3%). Respiratory rate and absence of the tail pinch response were monitored regularly. The scalp was shaved and rats were placed in the stereotax. A heating pad was used to prevent hypothermia. Lactated Ringer's solution ( $5 \text{ ml kg}^{-1}$  subcutaneous), buprenorphine ( $0.05 \text{ mg kg}^{-1}$  subcutaneous) and enrofloxacin ( $5 \text{ mg kg}^{-1}$ , subcutaneous) were administered to rats. A midline incision was made to expose the skull and a craniotomy was performed unilaterally over the mPFC. Virus injections were delivered with a 10 μl syringe and 33 gauge bevelled needle injected at  $100 \text{ nl min}^{-1}$  using an injection pump. A single 1 μl injection (AAV5-CKIIa-SSFO-eYFP at a titre of  $2\text{--}4 \times 10^{12}$  particles per ml) was made at the following stereotaxic coordinate to target the vmPFC (infralimbic cortex): anteroposterior, +2.7 mm; mediolateral, 0.5 mm; dorsoventral, 5.0 mm from bregma. Following injection, the injection needle was kept at the injection site for 10 min then slowly withdrawn. The skin was sutured and the rats recovered in a clean cage under a heated lamp. Rats were euthanized and perfused 5 months following surgery to allow strong YFP expression in vmPFC terminals in the amygdala.

**Rabies tracing.** For rabies tracing experiments, 1 μl of stock rabies-ΔG-GFP solution (provided by B. K. Lim and R. C. Malenka) was injected in the BMA (anteroposterior, –1.3; mediolateral, 2.7; dorsoventral, 5.6). Mice were sacrificed 6 days following surgery. Brain slices were then obtained as described below in 'Histology'. GFP-expressing cells were then counted on a Leica TCS SP5 scanning laser confocal microscope. Values plotted represent averages from 12 brain slices (three slices from four different mice).

**Retrobead tracing.** Red retrobeads (LumaFluor) were injected undiluted using a 10 μl nanofill syringe (World Precision Instruments) at  $0.1 \mu\text{l min}^{-1}$ . The syringe was coupled to a 33 gauge bevelled needle, and the bevel was placed to face the anterior side of the animal. The syringe was slowly retracted 20 min after the start of the infusion.

**Light delivery.** For Chr2 mice, blue light was generated by a 473 nm laser (Omicron Laserage) at 10 mW power to stimulate fibres in the amygdala and at 1 mW to activate mPFC cell bodies. Yellow light was generated by a 593.5 nm DPSS laser (MGL-F593.5; OEM Laser Systems), and bilaterally delivered to mice at 10 mW. A Master-8 pulse generator (A.M.P.I.) was used to drive the laser at 10 Hz, with 5-millisecond pulses for the entire duration of the tone (20 s). The laser output was delivered to the animal via an optical fibre (200 μm core, 0.22 numerical aperture, Doric Lenses), which was coupled to the fibre optic implanted on the animals through a zirconia sleeve.

**Elevated plus maze (EPM) and open field testing.** Experiments were performed 3 weeks after surgery for vmPFC:Chr2 and BMA:Chr2 groups and 3 months after surgery for vmPFC-amy:Chr2, vmPFC-amy:NpHR and dmPFC-amy:Chr2 groups, to allow for sufficient opsin expression. Mice were handled for 3 days before behavioural testing. On the day of the experiment, the fibre optic cannulae implanted on animals were connected to a patch cord. The animal was allowed to recover from this handling for 1–5 min in a cage before behavioural testing. The elevated plus maze was made of plastic painted grey (Med Associates), and the open field was a custom-made box with white plastic walls (dimensions:  $50 \times 50 \times 50 \text{ cm}$ ). In the beginning of the test mice were gently placed in the periphery of the open field or the closed arms of the EPM. The EPM and open field sessions lasted 15 and 20 min, respectively. The only exception is for the



experiment with corticosterone-treated mice, which explored the EPM for only 5 min. For both paradigms the session was divided into blocks of 5 min, which started with an 'off' epoch and then alternated between 'off' and 'on' epochs. Laser light was delivered for 5-min during 'on' epochs only. Both opsin and YFP-expressing animals that spent less than 5% time exploring the open arms of the EPM or centre of the open field were excluded from assays in which the optogenetic manipulation was expected to increase anxiety, as these animals already had ceiling levels of anxiety at baseline. This exclusion criterion was pre-established before the start of the experiment. In the experiment in which mice were treated chronically with corticosterone, mice ran the EPM for only 5 min, and blue light was delivered during the entire session to all groups of mice as 10-Hz trains of 5-ms pulses. For all anxiety assays animal position tracking, behavioural scoring and velocity measurements were done by the Biobase software automatically. All plots showing overall velocity were from open field exploration sessions. The experimenter was blind to the opsin group of the animal while running the experiment.

**Chronic treatment with corticosterone.** Mice were anaesthetized with 1.5–3.0% isoflurane and a small incision was made on the side of the neck. Corticosterone or placebo pellets (#G-111, Innovative Research of America) were placed in the neck incision, and were positioned between the skin and the underlying muscle tissue. Behavioural experiments were performed 15 days following implantation of the pellets. We tested pellets containing 7.5 and 15 mg of corticosterone, which are active over 21 days. For a 30 g mouse these pellets correspond to doses of 12.5 and 25 mg kg<sup>-1</sup> corticosterone per day. The pellets are made of a biodegradable matrix of cholesterol and cellulose, and allow for constant release of corticosterone.

**Fear conditioning.** All mice were handled for three days before behavioural assays for 5 min per day. The pulse-generator driving the laser was synchronized to tone onset by coupling the laser pulse-generator to the tone with a TTL pulse originating from the fear conditioning software (FreezeFrame, Actimetrics). For cued fear conditioning experiments, mice were placed in a chamber with a grid floor connected to a shock generator (Coulbourn). Two minutes after being placed in the chamber, mice were exposed to 20-s tones (2.9 kHz, 85 dB) co-terminating with a foot shock at pseudo-random inter-trial intervals (~2 min average). The protocol in Fig. 2b used six tone-shock pairings (0.7 mA, 2-s shocks), and the experiments in Fig. 2a used four tone-shock pairings (0.4 mA, 1-s shocks). The chamber was cleaned with 70% ethanol at the end of each trial. The next day (day 2, fear extinction), mice were placed in a different context (floor and walls were changed and white vinegar was sprayed in the chamber). First, mice were exposed to one tone without light delivery, to measure fear retrieval (tone 1 on day 2). Subsequently, mice were exposed to 10 tones (tones 2–11 on day 2) to undergo fear extinction with paired blue light delivery (10 Hz train of 5-ms pulses for 20 s). On day 3 mice were placed in the same environment as in day 2 and were exposed to one presentation of the tone. For the strong contextual fear conditioning protocol shown in Fig. 2c, d, mice explored the fear conditioning chamber for 2 min. Subsequently, mice received five 1-s foot shocks (0.4 mA) with ~2 min intervals on day 1. On day 2 blue light was delivered bilaterally to the amygdala while mice explored the same environment and on day 3 mice were tested for extinction retrieval in the same environment. The mouse explored the environment for 5 min during days 2 and 3. The weaker contextual conditioning protocol shown in Fig. 2e, f was similar, except that mice were conditioned to three 0.4 mA 1-s shocks and exposed to the training environment for 20 min on day 2 with blue light delivered for the entire 20 min. The experimenter was blind to the opsin group of the animal while running the experiment.

**Respiratory and heart rate measurements.** Respiratory rate and heart rate were measured using a pulse oximeter (MouseOx Plus; Starr Life Sciences). Pulse oximetry works by placing a collar clip around the mouse's shaved neck. One side of the collar emits red and infrared light, which pass through the neck and is measured on the other side of the neck by a detector. Oxygen-bound haemoglobin and de-oxygenated haemoglobin respectively absorb the infrared and red light preferentially. The ratio of these absorbances can be used to calculate oxygen saturation. Additionally, the small dilations in blood vessels induced by heart beats and respiration can also be detected by the change in the absorbances of these wavelengths. The functioning and limitations of this technique have been described in detail elsewhere<sup>49</sup>.

Data was collected on a computer running MouseOx Plus software. Mice were shaved around the neck and acclimated to moving with the collar sensor used to by the pulse oximeter for three days. Additionally, mice were handled for three days before experimenting. Respiratory rate was recorded as a moving average of ten measurements recorded at 1 Hz. Heart rate was recorded as a moving average of five heart beats. Baseline respiratory rate was typically in the 160 to 230 breaths per min range, while heart rate was in the 690 to 790 beats per min

range. The MouseOx Plus software generates errors during abrupt movements made by the animal, as these motion artefacts generate unreliable readings. Only error-free data was used from sessions in which the recordings had at least 30% of error-free samples. Recordings were obtained for either 15 min only in the home cage or for 5 min in the home cage and 10 min in the open field. The experimenter was blind to the opsin group of the animal while running the experiment.

**Microdrive construction and implantation.** Microdrives were custom built using Neuralynx (Neuralynx) a 16-channel electronic interface board (EIB-16), as described elsewhere<sup>34</sup>. Stereotrodes were made with 25- $\mu$ m Formvar-coated tungsten microwires (M165260, California Fine Wires) and were inserted into a metal cannula (GHX-24, Component Supply Co.), which in turn was attached to the 'R' channel slot of the EIB-16. A fibre optic cannula was glued to the outer part of the metal cannula, and glued to the protruding shaft of the stereotrode bundle, with care being taken not to glue the electrode tips. The stereotrodes were cut so that they extended 0.5 mm beyond than the end of the tip of the fibre optic. This design allows for light delivered through the cannula to reach the cells recorded by the stereotrode tips without having the fibre optic damage the cells being recorded. The EIB-16 was secured with screws (SHCX-080-6; Small Parts) to a custom-built teflon platform. The platform and the EIB were in turn secured to teflon cuffs with screws, allowing the assembly containing the stereotrodes, EIB and fibre optic to be lowered by turning the screws after the teflon cuffs have been cemented to the animal's skull as described previously<sup>34</sup>. Before stereotactically implanting the microdrive, one screw was placed in the front of the skull to increase the cemented (Grip Dental Cement; Dentsply) assembly's physical stability. Another screw was placed over the cerebellum to serve as an electrical ground and was soldered to the ground slot in the EIB (slot labelled 'G').

**In vivo electrophysiology data acquisition.** Recordings were performed as described previously<sup>34</sup>. Mice recuperated from the implantation of the microdrive for one week. Animals were then handled for three days and habituated to moving with the tether connecting the microdrive to the recording system. Stereotrodes were advanced until at least four well-isolated single units were present. Mice were then placed in the EPM for 15 min. Recordings were made with a unity-gain head-stage preamplifier (HS-16, Neuralynx) coupled to a fine-wire cable. Spikes larger than 25  $\mu$ V were band-pass filtered (600–6,000 Hz) and recorded at 32 kHz. Neural data were acquired on a personal computer running Neuralynx Cheetah software.

Animal position was tracked at 30 Hz using an overhead camera that detected two small LEDs attached to the head-stage. In experiments that used blue light to activate vmPFC terminals in the amygdala, the output of the laser pulse generator was recorded along with local spiking activity by the same recording amplifiers, allowing synchronous acquisition of neural data and laser onset and offset times. Blue light was delivered using the same parameters employed in other behavioural assays described above (10-Hz trains of 5-ms pulses at 10 mW power). At the end of the experiment current was passed through one of the stereotrode channels to produce an electrical lesion in the site of recording. Recordings were performed for 35 min. During minutes 0 to 5 recordings were obtained in the home cage. From minutes 5 to 25 the animal explored the elevated plus maze (EPM). Lastly, during minutes 25 to 35 the animal explored the light/dark test (LDT).

**Antidromic spike recording.** Mice expressing ChR2 in vmPFC cell bodies (injections as for behaviour) were anaesthetized using 1–2% isoflurane. Stimulation light pulses were delivered to the BMA using the same optical fibre type, coordinates, and stimulation pulse parameters as in behavioural experiments (5-ms pulses at 10 Hz). We used higher light power (80 mW) to maximize the chance of antidromic activation. Multiunit activity was simultaneously recorded across 32 in 2 animals (total 64 sites) spanning the entire mPFC (mediolateral, 0.4 mm; anteroposterior, 1.6 mm; dorsoventral, 1.8–3.3 mm from bregma) using a silicon probe (A1x32-6mm-50-177-Z32, NeuroNexus). Signals from each recording site were amplified, digitized and digitally filtered (600–6,000 Hz) using a multichannel recording system (RZ5D, Tucker-Davis Technologies). Multiunit firing was assessed by thresholding raw traces at three standard deviations from baseline using custom analysis software (MATLAB, Mathworks). As a positive control, the stimulation optical fibre was retargeted to the recording location in mPFC using an angled approach to directly elicit local spiking.

**Analysis of neural data.** *Code availability.* All custom-written MATLAB code is available upon request.

**Elevated plus maze.** Following recording, single units were clustered manually offline with the SpikeSort3D software (Neuralynx). Data were then imported into MATLAB using custom-written scripts. Peri-stimulus time histograms (PSTHs) of spiking activity were constructed for a 100 ms time window centred at the onset of 5-ms laser pulses. EPM scores were computed for all recorded single units to

estimate the extent to which each unit encoded arm type in the EPM<sup>24,34</sup>. The calculation of EPM scores has been explained in great detail in step-by-step fashion previously<sup>24</sup>. In brief, fold-firing rates relative to the unit's mean firing rate were calculated for each of the EPM's compartments, and the score was computed by the formula

EPM Score =  $(A - B)/(A + B)$ , where

$A = 0.25 \times (|FL - FU| + |FL - FD| + |FR - FU| + |FR - FD|)$  and

$B = 0.5 \times (|FL - FR| + |FU - FD|)$

FL, FR, FU and FD are the fold-firing rate in left, right, up and down arms, respectively. For example, if the mean firing rate in the whole maze is 4 Hz, and the unit fired at 8 Hz in the left arm, then  $FL = 8/4 = 2$ . Left and right arms are closed arms, and up and down arms are open arms.

'A' is the mean difference in normalized firing rate between arms of different types, while 'B' is the mean difference for arms of the same type. The maximum score of 1 indicates that all differences in firing between different arms can be explained by arm type ( $B = 0$ , where a unit has identical high rates only closed arms or only in open arms). A high EPM score indicates that a unit fires preferentially either in both closed arms or in both open arms. A score of zero indicates that a single unit has the same firing rate in all compartments of the EPM. A negative score would be assigned to a single unit that has similar firing rates in arms of different types (for example, a unit has high rates in one of the open arms and one of the closed arms).

Cells with EPM scores equal or smaller than zero were classified as not task-related. The remaining cells (all of which have EPM scores  $>0$ ) were classified as either closed- or open-arm-preferring, if the cell displayed higher rates in the closed or open arms, respectively.

To calculate if the population of experimentally observed EPM scores was significantly different than expected by chance, a simulated distribution of scores was generated. For each unit with  $n$  spikes, 500 simulated scores were generated by calculating the EPM score of  $n$  randomly chosen timestamps 500 times. This artificially generated population of EPM scores was then compared to the scores obtained from actual BMA single units using Wilcoxon's Rank sum test.

In order to analyse the stability of firing patterns of BMA cells in the EPM we calculated firing rates in the closed and open arms for the first and second halves of the EPM exploration session separately. We only used data from cells that fired at least 10 spikes in each arm type in each of the two halves of the session. EPM scores and preference of firing in a specific arm type were not significantly correlated with action potential waveform features that are commonly used to classify cells as excitatory or inhibitory (Supplementary Note 5).

**Cued fear conditioning.** Cells were classified as tone responsive if their firing rate during the tone presentation period was significantly different ( $P < 0.05$  with the Wilcoxon rank sum test) from firing during baseline (the 40 s preceding the tone presentation).

**In vitro electrophysiology.** Three months after injections, acute 300- $\mu$ m coronal slices were prepared by transcardially perfusing the mice with an ice-cold sucrose solution (containing in mM: 125 NaCl, 2.5 KCl, 0.1 CaCl<sub>2</sub>, 3.9 MgCl<sub>2</sub>, 26 NaHCO<sub>3</sub>, 1.25 NaH<sub>2</sub>PO<sub>4</sub>-H<sub>2</sub>O, 2.5 glucose, 50 sucrose) and slicing the brain tissue in the same ice-cold sucrose solution using a vibratome (VT1200S, Leica). Slices containing the PFC were fixed in 4% paraformaldehyde and saved for verification of the ChR2 injection site. Slices containing the amygdala were allowed to recover for 1 h at 33 °C in artificial cerebrospinal fluid (aCSF; containing in mM: 125 NaCl, 2.5 KCl, 2 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 26 NaHCO<sub>3</sub>, 1.25 NaH<sub>2</sub>PO<sub>4</sub>-H<sub>2</sub>O, 11 glucose) bubbled with 95% O<sub>2</sub>/5% CO<sub>2</sub>. Whole-cell patch-clamp recordings were performed in the same aCSF solution at 30–32 °C. Where indicated, TTX (1  $\mu$ M) and 4-AP (100  $\mu$ M) were added to the aCSF. Resistance of the patch pipettes was 2.5–4 M $\Omega$  when filled with intracellular solution containing the following (in mM): 120 CsMeSO<sub>3</sub>, 15 CsCl, 8 NaCl, 0.2 EGTA, 10 HEPES, 2 Mg-ATP, 0.3 Na-GTP, 10 TEA (tetraethylammonium), 5 QX-314 (lidocaine *N*-ethyl bromide), adjusted to pH 7.25 with CsOH. In some experiments, 0.2% biocytin was also included in the intracellular solution. Slices containing biocytin-filled cells were later processed as explained in 'Immunocytochemistry, Biocytin', to allow accurate visualization of the recorded cell's location. Signals were amplified with a Multiclamp 700B amplifier, acquired using a Digidata 1440A digitizer, sampled at 10 kHz, and filtered at 2 kHz. All data acquisition and analysis were performed using pCLAMP software (Molecular Devices). Neurons were visually identified for patching using an upright microscope (Olympus BX51WI) equipped with DIC optics, filter sets for visualizing YFP and tdtomato, and a CCD camera (RoleraXR, Q-Imaging). After break-in, neurons were voltage-clamped at  $-65$  mV. As neurons were recorded, pictures of the placement of the recording electrode were taken under low magnification ( $5\times/0.15$  air objective) and the location of the recording was noted on a schematic of the amygdala. To stimulate ChR2 expressed in axon terminals from

the PFC, 5-ms blue light pulses ( $\sim 10$  mW mm<sup>-2</sup>) were generated using a Spectra X LED light engine (Lumencor) and delivered to the slice via a  $40\times/0.8$  water-immersion objective focused onto the recorded neuron. Pulses were delivered once every 30 s. Responses sizes were calculated by baseline-subtracting and averaging 5–15 traces together, then calculating the peak amplitude in a 50 ms window after the light pulse. Neurons that did not show a peak amplitude in this window that exceeded 5 s.d. of the baseline noise were counted as non-responders. Recordings shown in Fig. 4 were performed in mice injected with 1  $\mu$ l of  $2 \times 10^{12}$  titre AAV5-K2aCamK2a-ChR2-YFP virus, similar to the infusions done in vmPFC-amygdala:ChR2 mice. Infusions done in GAD2-Cre mice were done with  $4 \times 10^{12}$  titre (Extended Data Fig. 7g). This difference is probably why a higher fraction of BMA neurons were responsive to optical stimulation of vmPFC terminal in the latter experiment (91%) compared to Fig. 4 (63%). Importantly, cells from all three amygdala regions studied (BMA, BLA and ITC) were recorded from each mouse. This approach ensures that the unresponsiveness of ITC and BLA cells in mice expressing ChR2 in the vmPFC is not due to insufficient opsin expression or any other technical problem, as responsive BMA cells were recorded from the same animal.

**Histology.** Mice were deeply anaesthetized and transcardially perfused with ice-cold 4% paraformaldehyde in PBS. Brains were fixed overnight in 4% paraformaldehyde and then equilibrated in 30% sucrose in PBS. Brains were sliced in 40- $\mu$ m-thick coronal slices in a freezing microtome, and stored in cryoprotectant solution (a mixture of 5:6:9 volumes of glycerol, ethylene glycol and PBS) at 4 °C until being mounted on slides on PVA-DABCO. Nuclei were stained by incubating with DAPI (1:50,000 in PBS, 25 min). In animals with retrobead injections brain sections were stored in PBS at 4 °C in Fluoromount-G (Southern Biotech). Confocal images were obtained on a Leica TCS SP5 scanning laser microscope using a  $20\times/0.70$  numerical aperture oil immersion objective. Only mice with opsin expression restricted to the target region and correct fibre optic cannula placement were used.

**Immunocytochemistry and c-Fos counting: c-Fos.** vmPFC-amygdala:ChR2 mice received blue light in the amygdala (10 Hz, 5-ms pulses, 10 mW power) for 10 min. Animals were perfused 90 min after receiving optical stimulation. Following perfusion, brain slices were obtained as described above in 'Histology'. c-Fos staining was performed as described previously<sup>46</sup>. In brief, coronal sections containing the amygdala were washed in PBS (three 10-min washes), and were then incubated for 1 h in blocking solution with 0.2% Triton-X-100 and 2% normal donkey serum. Sections were then incubated overnight with anti c-Fos primary antibody (rabbit anti c-Fos 1:500, Cell Signaling Technology, cat. no. 2250S). The next day the sections were washed in PBS (three 10-min washes) and incubated at 1 h at room temperature with the secondary antibody at 1:500 dilution (donkey anti-rabbit antibody conjugated to Cy3, Jackson Laboratories). Slices were washed three times in PBS and mounted on slides with PVA-Dabco (Sigma). c-Fos counting was performed blinded to treatment group on a Leica TCS SP5 scanning laser confocal microscope. To stain cells against GABA the same procedure was employed, but this time using anti-GABA antibody (Sigma, cat. no. A2052).

**Immunocytochemistry and c-Fos counting: FoxP2.** Staining against FoxP2 was done similarly to the procedure described above for c-Fos, but with the alterations described below. The first incubation was in 0.3% Triton-X in 3% normal donkey serum, and it lasted 2 h. The primary antibody used was anti-FoxP2 made in Rabbit (AbCam, cat. no. 16046) at 1:500 dilution for 12 h at 4 °C. The secondary antibody was anti-rabbit conjugated to Alexa Fluor 647 (AbCam, cat. no. 150075), incubated at 1:500 dilution for 3 h at room temperature.

**Immunocytochemistry and c-Fos counting: GABA.** Same procedure as FoxP2 staining, except that anti-GABA antibody was used (Sigma, cat. no. A2052), with 0.1% Triton-X.

**Immunocytochemistry and c-Fos counting: biocytin.** The same procedure used to stain against c-Fos was employed to visualize cells filled with biocytin during patch clamping recordings, with the following modifications. The first incubation was in 0.3% Triton-X, and 3% normal donkey serum, and it lasted for 2 h. Instead of an incubation with a secondary antibody sections were incubated with streptavidin conjugated to Alexa Fluor 647 (5  $\mu$ g ml<sup>-1</sup>) for 2 h at room temperature.

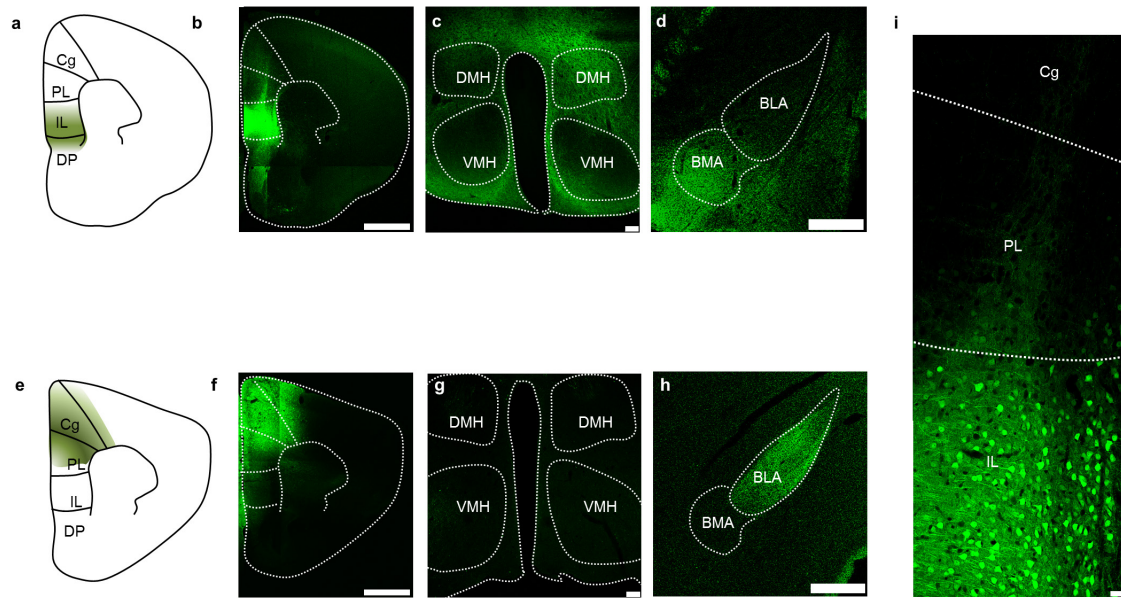
**CLARITY.** Slices from vmPFC-amygdala:YFP mice were clarified using the CLARITY procedure as explained elsewhere<sup>50</sup>. In brief, after patching, 300- $\mu$ m sections were fixed in hydrogel solution (4% PFA, 1% acrylamide w/ Bis) for 20 h at 4 °C. After polymerization (37 °C, 4 h), the sections were clarified in 4% SDS for 2 days (at 37 °C). The cleared sections were washed three times with PBST (0.1% TritonX) for a total of 4 h at room temperature. Alexa647-conjugated anti-GFP antibody (1:100, Invitrogen) was used to stain YFP (room temperature, overnight). Stained sections were incubated in FocusClear (CellExplorer Labs Co.) for 1 h before imaging with a confocal microscope.



**Statistics.** Two-way repeated measures ANOVA was used. Wilcoxon rank sum post hoc tests were used following the ANOVA only if significant main or interaction effects were detected. Two-tailed tests were used throughout with  $\alpha = 0.05$ . The Wilcoxon test is non-parametric, and as such it does not require the data to be normally distributed. Nevertheless, we tested the normality of all the behavioural data being used with the Lilliefors test, and found that our behavioural data passed this test. The variances of the groups being compared were not significantly different in any of the comparisons. Variance equality was tested using the *F*-test. Asterisks in the figures indicate the *P* values for the post hoc test at a given epoch. Standard error of the mean (s.e.m.) was plotted in each figure as an estimate of variation within each group of data. Multiple comparisons were adjusted with the false discovery rate method. Our experimental design of having light stimulation 'on' and 'off' epochs in a single session for each animal allows us

to use repeated measures ANOVA to test if light delivery significantly alters the behaviour of interest relative to its own baseline (that is, the first light off epoch). This strategy is statistically more powerful than comparing YFP and ChR2 groups with an unpaired test. As this statistical test evaluates changes relative to baseline, we opted to plot normalized data. In this way the data presentation accurately represents the differences that are being statistically tested. Importantly, there were no significant differences at baseline between YFP and experimental groups in behavioural or physiological measures of anxiety. In all figure legends *n* refers to biological replicates.

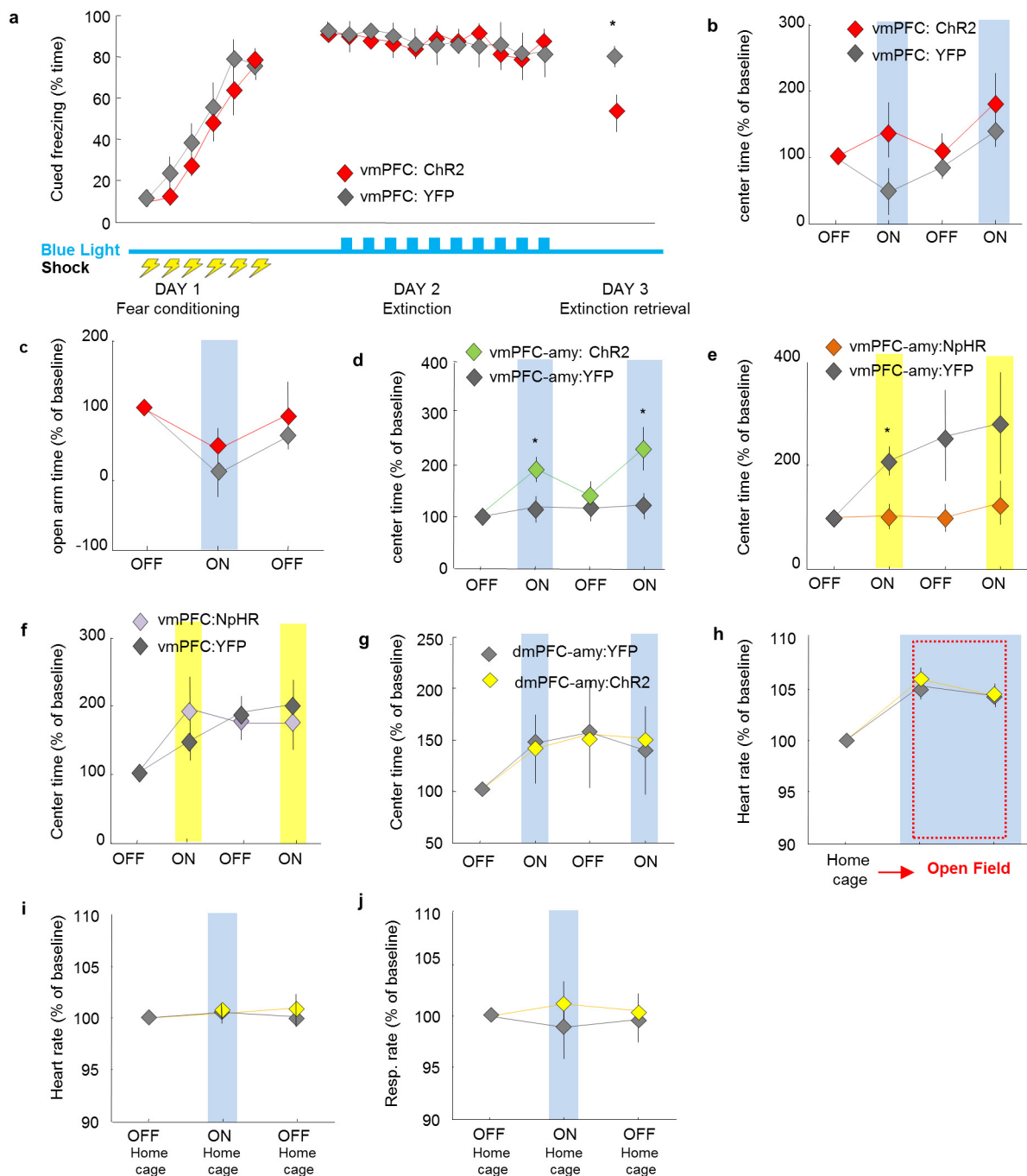
49. Smith, J. Understanding pulse oximetry. *Anaesth. Intensive Care* **20**, 255–256 (1992).
50. Torner, R., Ye, L., Hsueh, B. & Deisseroth, K. Advanced CLARITY for rapid and high-resolution imaging of intact tissues. *Nature Protocols* **9**, 1682–1697 (2014).



**Extended Data Figure 1 | Methodology necessary for targeting viral infusions restricted to vmPFC and dmPFC.** **a**, Scheme showing subregions of the mPFC. **b**, Representative example from one mouse showing a coronal section containing the vmPFC (IL+DP) in a mouse expressing YFP in the vmPFC. **c**, **d**, Coronal section depicting vmPFC fibres in the hypothalamus (**c**) and amygdala (**d**).  $n$  = example from 1 mouse chosen from  $n = 7$  mice (**a-d**). **e-h**, Same as **a-d**, but for an animal that received viral injection in the dmPFC (PL and Cg).  $n$  = example from

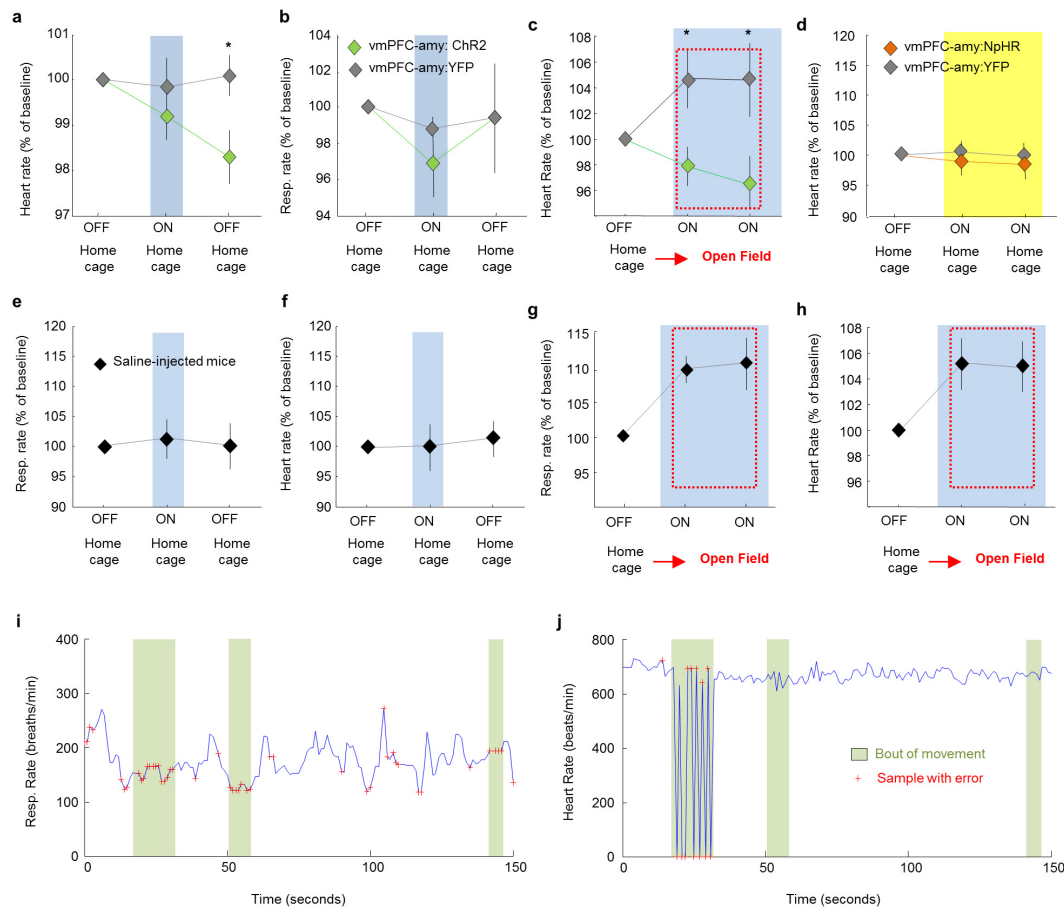
1 mouse chosen from  $n = 7$  mice (**e-h**). **c**, **g**, Note the presence of fibres from the vmPFC, but not dmPFC, in the hypothalamus. Specifically, the vmPFC projects strongly to the dorsomedial, but not ventromedial hypothalamus (DMH and VMH, respectively). **g**, Section showing expression of YFP in a vmPFC:YFP mouse.  $n$  = example from 1 mouse chosen from  $n = 7$  mice. Cg, cingulate cortex; PL, prelimbic cortex; IL, infralimbic cortex; DP, dorsal peduncular cortex. Scale bars, 1 mm (**b**, **f**); 100  $\mu$ m (**c**, **g**); 500  $\mu$ m (**d**, **h**); 35  $\mu$ m (**i**).





**Extended Data Figure 2 | Optogenetic manipulations involving the mPFC and its projections to the amygdala in fear and anxiety paradigms.** **a**, Mice were fear conditioned to six tone-shock (0.7 mA, 2-s shocks) pairings on day 1 (fear acquisition). On day 2 (fear extinction) animals were exposed to the tone in a different context for 11 trials. Blue light was delivered only on day 2, for trials 2–11. vmPFC:ChR2 mice froze less than control mice during day 3 (extinction retrieval). **b**, **c**, Blue light delivery did not change avoidance of open spaces in vmPFC:ChR2 mice relative to control mice in the open field (**b**) or the EPM (**c**).  $n = 7$  vmPFC:ChR2 and 7 vmPFC:YFP mice (**a–c**). **d**, Excitation of the vmPFC-amygdala projection with blue light increased exploration of the centre of the open field in vmPFC-amygdala:ChR2 mice relative to controls. Two-way repeated measures ANOVA, opsin  $\times$  epoch interaction,  $F_{3,68} = 3.1$ ,  $P = 0.03$ , post hoc Wilcoxon rank sum test  $P = 0.002$ .  $n = 12$  vmPFC-amygdala:ChR2 and 13 vmPFC-amygdala:YFP mice.

**e**, Inhibition of the same projection in vmPFC-amygdala:NpHR mice with yellow light increased avoidance of open spaces. Two-way repeated-measures ANOVA, main effect of opsin,  $F_{3,68} = 7.26$ ,  $P = 0.008$ , post-hoc Wilcoxon rank sum test  $P = 0.03$ .  $n = 14$  vmPFC-amygdala:NpHR and 11 vmPFC-amygdala:YFP mice. **f**, Same as **e**, but for inhibition of vmPFC cell bodies.  $n = 11$  vmPFC:NpHR and 8 vmPFC:YFP mice. **g–j**, Optogenetic stimulation of the dmPFC-amygdala projection did not alter behaviour in the open field (**g**), heart rate in the open field (**h**), heart rate in the home cage (**i**), or respiration rates in the home cage (**j**). Blue light stimulation epochs are labelled ON and/or with a blue bar.  $n = 7$  dmPFC-amygdala:YFP and 7 dmPFC-amygdala:ChR2 mice. **b–j**, Data are plotted in 5-min consecutive intervals. Light delivery epochs are labelled ON and/or with a blue or yellow bar. \* $P < 0.05$ , Wilcoxon rank sum test; error bars,  $\pm$  s.e.m.;  $n$  refers to biological replicates.

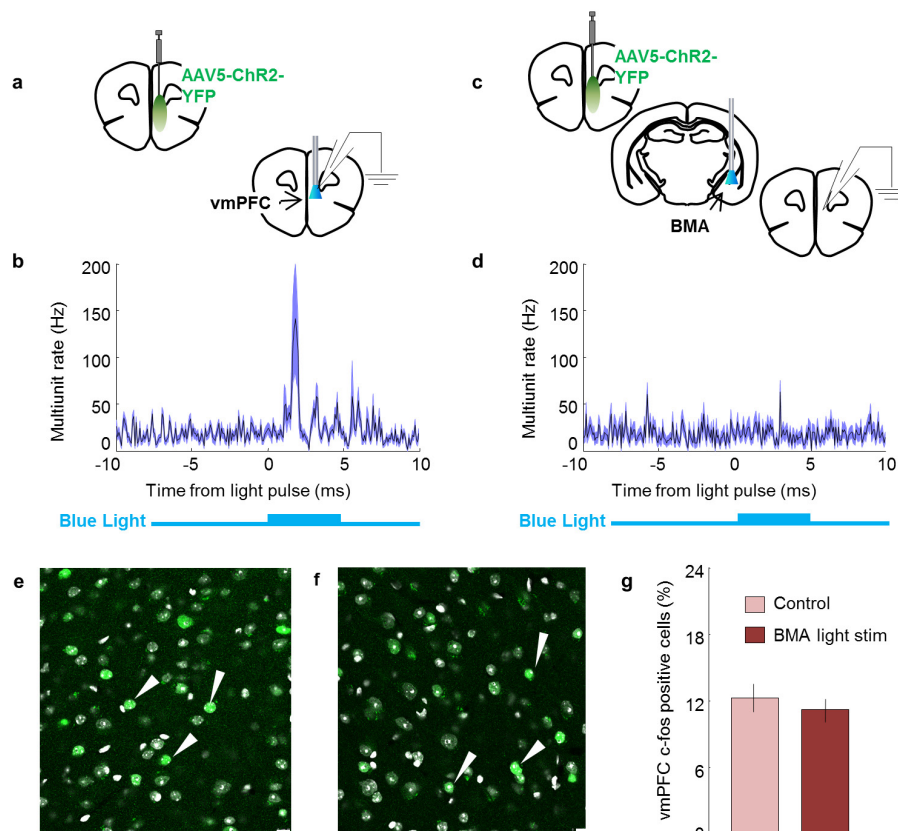


### Extended Data Figure 3 | Heart and respiratory rate changes elicited by optogenetic manipulation of vmPFC fibres in the amygdala.

**a, b**, Optogenetic stimulation of vmPFC fibres in the amygdala in the home cage did not significantly alter heart (**a**) or respiratory rate (**b**) in mice in the home cage during the light ON period. Nevertheless, a downward trend was observed for both measurements during delivery of blue light.  $n = 12$  vmPFC-amygdala:ChR2 and 6 vmPFC-amygdala:YFP mice. **c**, Blue light delivery in vmPFC-amygdala:ChR2, but not control, mice prevented increases in heart rate in the open field test (OFT) relative to the home cage. Two-way repeated measures ANOVA, main effect of opsin,  $F_{2,29} = 10.98$ ,  $P = 0.0019$ , post-hoc Wilcoxon rank sum test  $P = 0.04$ ;  $n = 11$  vmPFC-amygdala:ChR2 and 6 vmPFC-amygdala:YFP mice. **d**, Inhibition of the vmPFC-amygdala projection in vmPFC-amygdala:NpHR mice with yellow light in the home cage did not alter heart rate.  $n = 6$  vmPFC-amygdala:NpHR; 6 vmPFC-amygdala:YFP mice. **e–h**, Mice were injected with saline in the vmPFC. Fibre optics were placed above the BMA. **e, f**, Delivery of blue light did not alter respiratory rate (**e**)

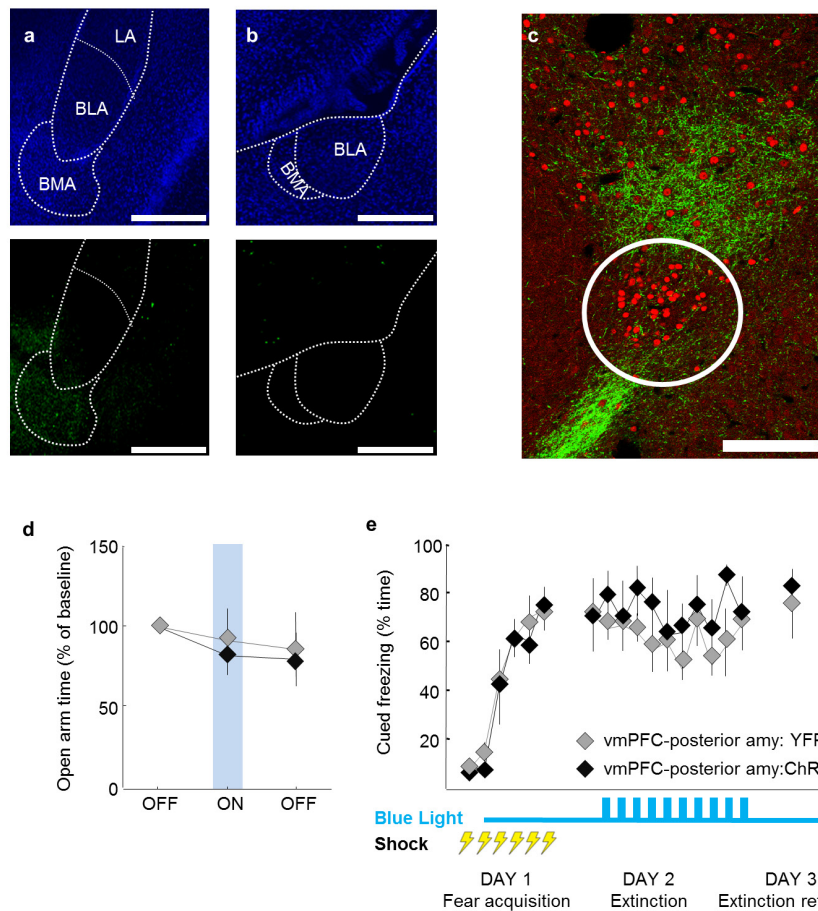
or heart rate (**f**) in the home cage. **g, h**, Respiratory rate (**g**) and heart rate (**h**) increased, relative to the home cage, when mice were placed in the anxiogenic open field. Blue light delivery did not prevent the increase in respiratory and heart rate observed in the open field.  $n = 7$  sham mice. **a–h**, Data are plotted in 5-min consecutive intervals. Light stimulation epochs are labelled with ON and with a blue or yellow bar. **i, j**, Example raw traces of respiratory (**i**) and heart rate (**j**) recorded at 1 Hz obtained from a freely moving mouse through pulse oximetry. Movement bouts are shown in green, and single samples with errors due to motion artefacts are shown as red crosses. Error samples are detected automatically by software (Starr Life Sciences). **i**, Most error samples occur during movement bouts and a few errors can be seen outside of movement bouts in the respiratory rate trace. **j**, Heart rate recordings are generally stable and errors occur only during prolonged and large movement bouts. Samples with errors were not used in any other plot or data analysis. Representative traces from one mouse. Error bars,  $\pm$  s.e.m.;  $n$  refers to biological replicates.





**Extended Data Figure 4 | Stimulation of ChR2-expressing vmPFC terminals in the basomedial amygdala: lack of detection of antidromic spikes in vmPFC.** **a**, Mice were injected with AAV5-CamK2 $\alpha$ -ChR2-YFP in the vmPFC. Blue light was delivered above the vmPFC. Simultaneous *in vivo* anaesthetized recordings under isoflurane were obtained from the vmPFC. **b**, Average of 64 recording sites in the mPFC showing that blue light elicited orthodromic spikes in ChR2-expressing cortical cells. **c**, Same as **a**, but blue light was delivered to ChR2-expressing vmPFC terminals in the BMA while recordings were obtained from the mPFC. **d**, Average of 64 recording mPFC sites showing that multiunit activity in the mPFC did not detectably increase following excitation of vmPFC terminals in the amygdala. Recordings with delivery of blue light to the vmPFC (**a**)

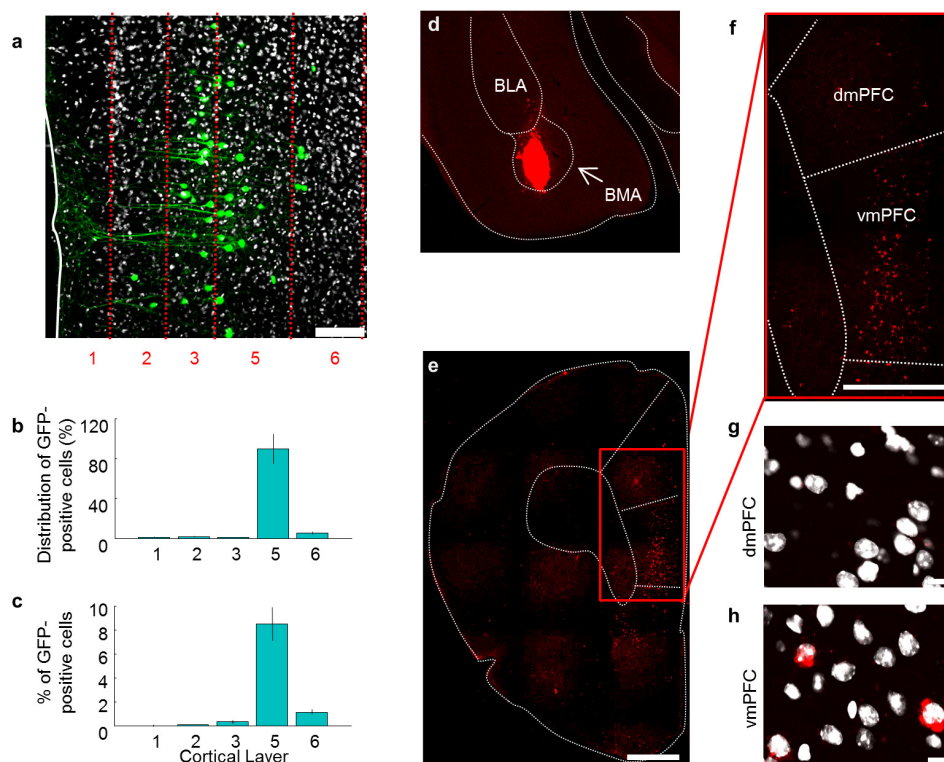
or BMA (**c**) were obtained from the same mice. The 5 ms blue light pulse is shown in blue below the graph. A 32-site recording electrode probe was used to target deep cortical layers.  $n = 64$  sites from 2 animals (**b**, **d**). **e–f**, Compared to baseline controls (**e**), stimulation of ChR2-expressing vmPFC fibres in the BMA of freely behaving awake animals (**f**) did not change c-Fos expression in deep layers of the vmPFC (layers 5 and 6). **g**, Summary bar graph showing the mean percentage of c-Fos positive cells in control animals and mice with stimulation of vmPFC fibres in the BMA.  $n = 5$  animals for each group. **e–f**, Arrowheads indicate examples of c-Fos-expressing cells. Scale bar, 10  $\mu$ m; error bars,  $\pm$ s.e.m.;  $n$  refers to biological replicates.



**Extended Data Figure 5 | vmPFC innervation of the amygdala in mice and rats.** **a, b,** Mice were injected with AAV5-CamK2 $\alpha$ -YFP in the vmPFC and fibres were imaged in the amygdala. **a,** Top, DAPI-stained amygdala section. Bottom, vmPFC fibres in the BMA 1.3 mm posterior from bregma. This coordinate was used for the fibre optic implantation in the vmPFC-amygdala behavioural cohorts. **b,** Same as **a**, but for a more posterior section (2.3 mm from bregma), showing no prominent vmPFC fibres of passage that traverse the BMA and terminate elsewhere. Nuclei were stained with DAPI,  $n = 4$  mice. Scale bar, 0.5 mm. **c,** Rats were injected with AAV5-CaMK2 $\alpha$ -SSFO-YFP in the vmPFC (infralimbic cortex). Six months following viral injection brain slices were stained for FoxP2 to identify ITCs (red). The representative image shows vmPFC fibres (green) surrounding an ITC cluster (circled in white). Note that

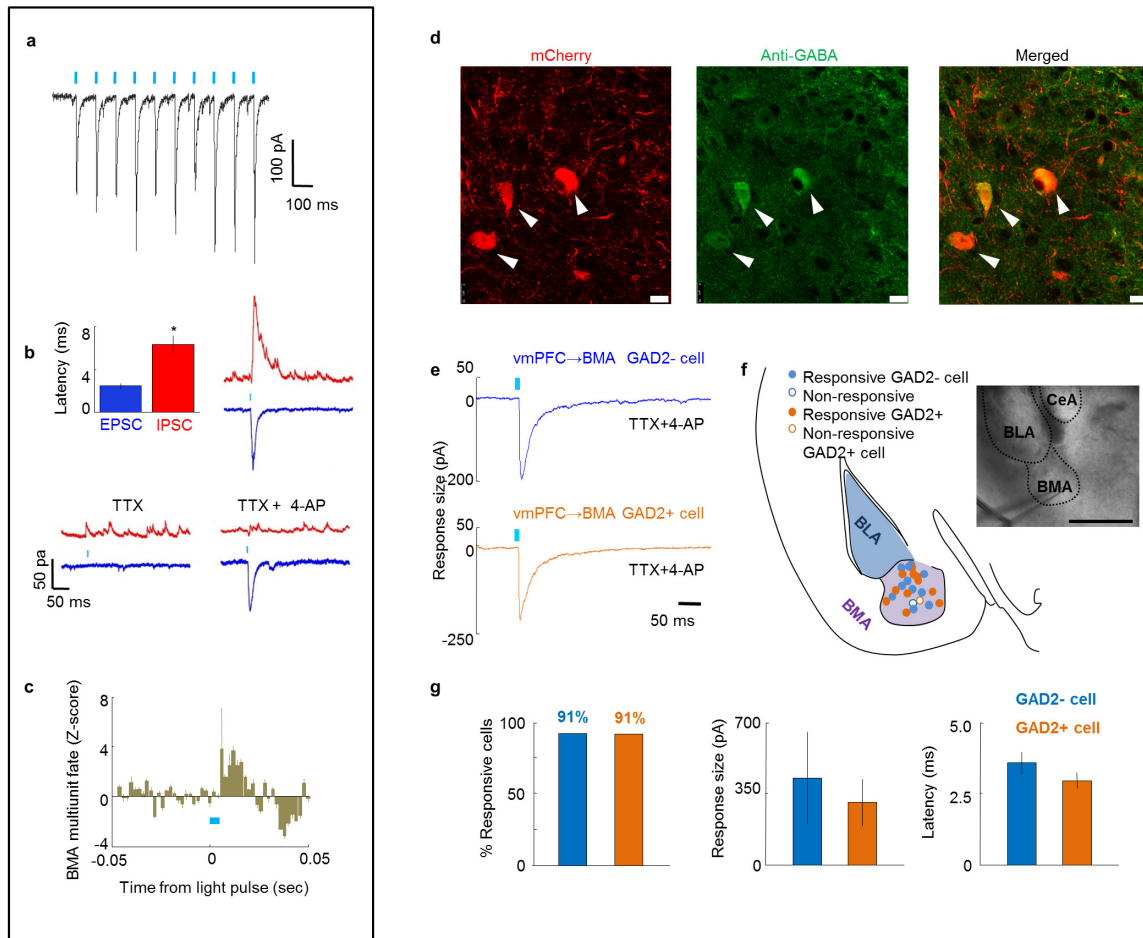
the vmPFC does not strongly innervate the ITCs in rats. Nevertheless, a sparse vmPFC-ITC projection can be seen. Image from one representative animal chosen from  $n = 3$  rats. Scale bar, 100  $\mu$ m. **d, e,** Mice were injected with AAV5-CamK2 $\alpha$ -ChR2-YFP in the vmPFC. Fibre optics were placed above the amygdala (amy), but 500  $\mu$ m posterior to the implants shown in Fig. 1. Delivery of blue light to this posterior amygdala site did not alter exploration of the open arms in the elevated plus maze (**d**) or freezing in cued fear conditioning (**e**), suggesting that activation of vmPFC fibres of passage that go beyond the amygdala do not have an important role in regulating anxiety and fear.  $n = 7$  vmPFC-posterior amygdala:YFP and 8 vmPFC-posterior amygdala:ChR2 mice. Error bars,  $\pm$  s.e.m.;  $n$  refers to biological replicates.





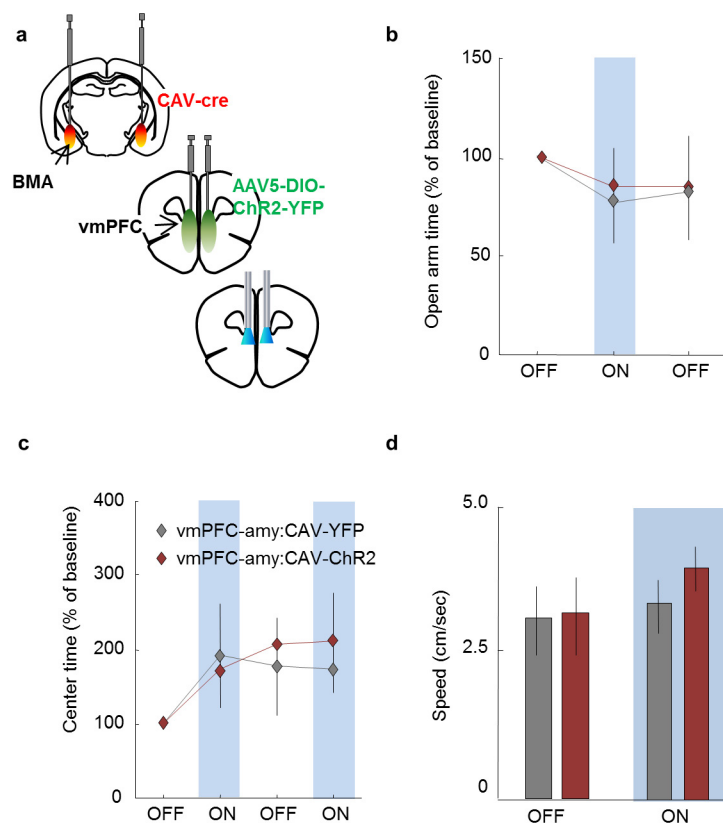
**Extended Data Figure 6 | Quantification of BMA-projecting vmPFC neurons.** **a–c**, Mice were injected with the retrogradely propagating  $\Delta$ G rabies-GFP virus in the basomedial amygdala (BMA). **a**, Ten days after viral infusion, retrogradely labelled vmPFC cells can be seen expressing GFP. The number of GFP-expressing vmPFC cells was quantified across layers, both as a percentage of all GFP-positive cells (**b**) and as a percentage of all vmPFC cells (**c**) (counting labelled and unlabelled cells).  $n = 4$  mice; scale bar, 75  $\mu$ m (**a**). **d**, Mice were injected with retrobeads in the BMA.

**e**, Image of a coronal section containing the mPFC. Note the presence of retrobead-containing cells in the vmPFC. **f**, Expanded image of the zone demarcated by a red rectangle in **e**. Labelled cells can be seen in the vmPFC, but not the dmPFC. **g, h**, Confocal image showing unlabelled cells in the dmPFC (**g**) and labelled cells in the vmPFC (**h**). **a, g, h**, Nuclei were stained with DAPI.  $n = 5$  mice (**d–h**). Scale bars, 250  $\mu$ m (**d, f**); 500  $\mu$ m (**e, f**); 10  $\mu$ m (**g, h**). Error bars,  $\pm$ s.e.m.;  $n$  refers to biological replicates.



**Extended Data Figure 7 | Characterization of the vmPFC-BMA projection by optical stimulation of vmPFC terminals *in vivo* and *in vitro*.** **a**, Example trace from one mouse showing responses in a BMA cell following a train of 5-ms 10 Hz pulses in an acute brain slice with ChR2-expressing vmPFC terminals. These were the same parameters used for behavioural optogenetic experiments. **b**, Optogenetic stimulation of vmPFC fibres in the BMA in acute brain slices elicited both IPSCs (red) and EPSCs (blue), which had significantly different latencies. TTX abolished both IPSCs and EPSCs. 4-AP was added in the presence of TTX to rescue monosynaptic responses. Note that 4-AP rescued the EPSC, but not the IPSC.  $n = 7$  cells from  $n = 2$  mice. **c**, BMA multiunit recordings were obtained in awake behaving mice during optical stimulation of ChR2-expressing vmPFC terminals. Activation of vmPFC terminals dramatically increased firing rates in the BMA. The graph shown is an average of  $n = 14$  multiunit recordings from  $n = 4$  mice. **d**, A GAD2-Cre mouse was injected with AAV5-DIO-mCherry in the BMA. First panel shows antibody staining against GABA. Middle panel shows expression of mCherry in Cre-expressing cells. Last panel shows a merged photo

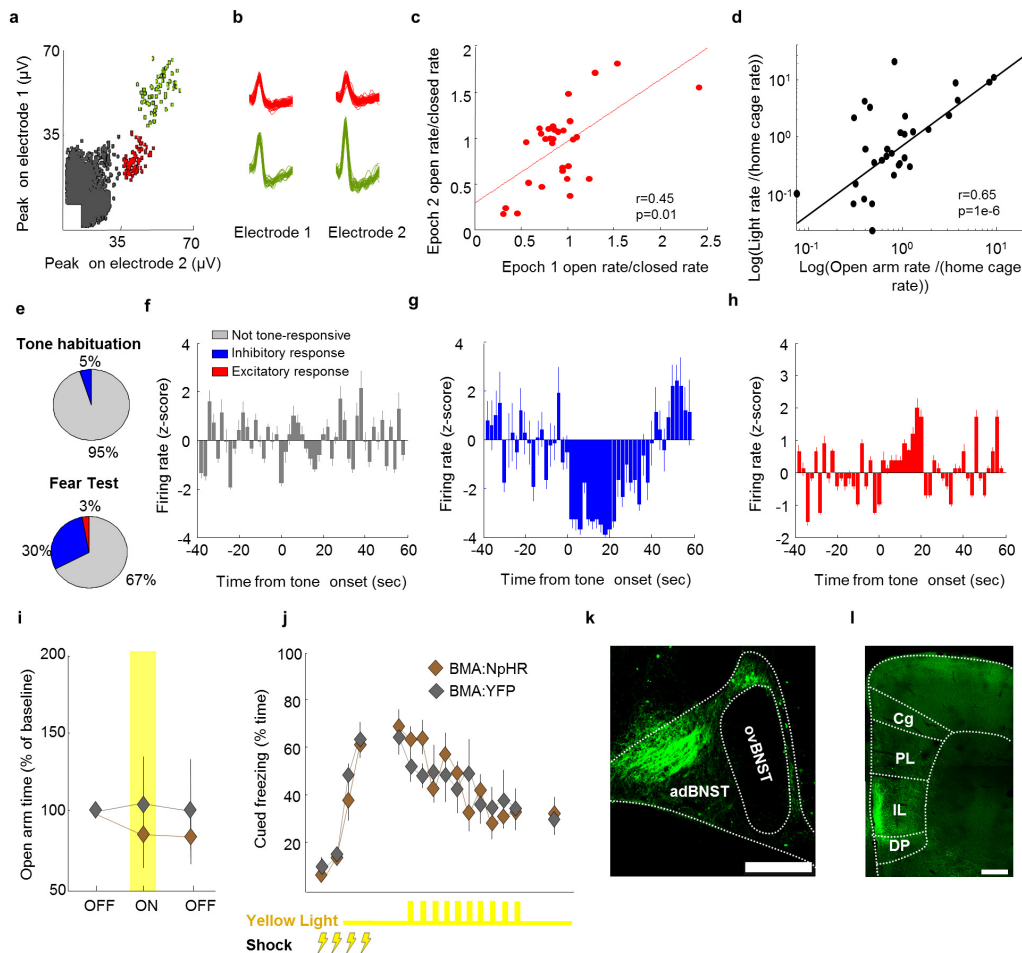
of the first two panels. Note overlap of mCherry expression and GABA staining. Arrowheads show examples of double-labelled cells.  $n = 5$  mice. **e**, Example traces from one mouse (chosen from  $n = 3$  mice) showing stimulation of ChR2-expressing vmPFC terminals in amygdala acute slices elicits responses in both GAD2 negative (putative excitatory projection cells) and positive cells (inhibitory interneurons). Recordings were done in the presence of TTX and 4-AP to abolish polysynaptic responses. **f**, Scheme displaying the location of all recorded cells. Responsive cells are shown as filled circles. Inset shows a BMA cell being patched. Inset scale bar, 10  $\mu\text{m}$ . **g**, Left: mean percentage of responsive cells. Middle: average response size of recorded cells. Right: Average latency of recorded responses relative to the start of the light pulse.,  $n = 12$  GAD2 positive and 12 GAD2 negative cells (from  $n = 3$  mice) (**e–g**). 4-AP, 4-aminopyridine; TTX, tetrodotoxin; IPSC, inhibitory postsynaptic current; EPSC, excitatory postsynaptic current. **a**, **b**, **c**, **e**, A 5 ms pulse of blue light (indicated by a blue tick mark) was used to elicit stimulation.  $*P < 0.05$ ; error bars,  $\pm$ s.e.m.;  $n$  refers to biological replicates.



**Extended Data Figure 8 | Activation of BMA-projecting vmPFC cells decreases cued fear.** **a**, Mice were injected with the retrogradely propagating canine adenovirus encoding Cre recombinase (CAV-Cre) in the BMA. Mice were also injected with a viral vector that induces expression of ChR2-YFP or YFP only in the presence of Cre recombinase (AAV5-DIO-ChR2-YFP). Mice were implanted bilaterally with fibre

optics above the vmPFC for delivery of blue light. **b–d**, Delivery of blue light (10 Hz, 5-ms pulses at 10 mW) did not alter exploration of the open arms in the EPM (**b**), the centre of the open field (**c**), or speed (**d**).  $n = 7$  vmPFC-amygdala:CAV-YFP and 8 vmPFC-amygdala:CAV-ChR2 mice. Error bars,  $\pm$  s.e.m.;  $n$  refers to biological replicates.





**Extended Data Figure 9 | BMA activity and function in anxiety and fear paradigms.** **a**, Example isolated BMA single-unit spike clusters recorded with stereotrodes *in vivo*. **b**, Waveforms of the single-unit clusters shown in **a**, as recorded on each of the two electrodes comprising the stereotrode. **c**, Ratios of BMA neuron open arm/closed arm firing rates are shown for minutes 1 to 10 (epoch 1) and minutes 10 to 20 (epoch 2) of a 20 min exploration session of the elevated plus maze (EPM). Open/closed firing rate ratios are highly correlated across both epochs ( $r = 0.45$ , Spearman correlation), indicating that BMA firing patterns were stable throughout the entire 20 min session. **d**, The same cells shown in **c** were also recorded in the home cage and in the light/dark test. Firing rates in the light compartment of the light/dark test and the open arms of the EPM (plotted as fold-increase of rates from the non-anxiogenic home cage) were highly correlated ( $r = 0.65$ , Spearman correlation), indicating that BMA neurons respond similarly to anxiety induced by two different anxiogenic stimuli (bright lights and open areas).  $n = 38$  cells from  $n = 4$  mice (**a–d**). **e–h**, Recordings were obtained from basomedial amygdala (BMA) cells during presentation of a fear conditioned auditory tone. **e**, Top, distribution of responsive cells to the auditory tone before fear

conditioning. Bottom, same as in upper panel, but for a fear recall test. The proportion of responsive cells increased following fear conditioning. Note that the vast majority of tone-responsive cells showed decreases in firing rate during the presentation of the fear-conditioned tone. **f**, Example cell that was not tone-responsive. **g**, **h**, Example cells that are inhibited (**g**) or excited (**h**) during tone presentation. **e**,  $n = 20$  cells during habituation and 71 cells during fear recall. **f–h**, Data are an average of ten tone presentations for each of the three cells shown.  $n = 4$  mice (**a–h**). **i**, Mice were injected with AAV5-CamK2 $\alpha$ -NpHR-YFP in the BMA. **i**, **j**, Yellow light didn't change behaviour in the elevated plus maze (**i**), or cued fear extinction (**j**).  $n = 8$  BMA:NpHR and 7 BMA:YFP mice (**i**, **j**). **k**, Eight weeks after viral injections BMA projections can be seen in BMA:YFP mice in the anterodorsal bed nucleus of the stria terminalis (adBNST) but not in the oval BNST (ovBNST). **l**, Prominent BMA innervation was also visible in the infralimbic cortex (IL), but not in the prelimbic (PL), dorsal peduncular (DP) or cingulate cortices (Cg). Images from one representative mouse chosen from  $n = 9$  BMA:YFP mice. Scale bars, 250  $\mu\text{m}$  (**a**, **b**); 500  $\mu\text{m}$  (**k**, **l**). Error bars,  $\pm$  s.e.m.;  $n$  refers to biological replicates.

Extended Data Table 1 | Anxiety behavioural data in absolute values

	Figure location	0-5 min.	5-10 min.	10-15 min.	15-20 min.
vmPFC-amy:YFP EPM	1b	57±12	73±15	63±19	
vmPFC-amy:ChR2 EPM	1b	63±14	50±11	61±14	
dmPFC-amy:YFP EPM	1f	53±7	48±17	38±5	
dmPFC-amy:ChR2 EPM	1f	50±12	46±16	42±17	
BMA:YFP EPM	5j	64±14	35±11	31±14	
BMA:ChR2 EPM	5j	65±19	77±23	45±19	
BMA:YFP OFT	5l	62±9.6	84±13	91±17	122±23
BMA:NpHR OFT	5l	66±8.5	66±10	73±13	71±9.5

Table displaying time spent (in seconds) in the open arms of the elevated plus maze (EPM) or the centre of the open field test (OFT). Values are expressed as averages ± s.e.m.

# Oxidative stress inhibits distant metastasis by human melanoma cells

Elena Piskounova<sup>1</sup>, Michalis Agathocleous<sup>1</sup>, Malea M. Murphy<sup>1</sup>, Zeping Hu<sup>1</sup>, Sara E. Huddleston<sup>1</sup>, Zhiyu Zhao<sup>1</sup>, A. Marilyn Leitch<sup>2</sup>, Timothy M. Johnson<sup>3</sup>, Ralph J. DeBerardinis<sup>1</sup> & Sean J. Morrison<sup>1,4</sup>

**Solid cancer cells commonly enter the blood and disseminate systemically, but are highly inefficient at forming distant metastases for poorly understood reasons. Here we studied human melanomas that differed in their metastasis histories in patients and in their capacity to metastasize in NOD-SCID-*Il2rg*<sup>-/-</sup> (NSG) mice. We show that melanomas had high frequencies of cells that formed subcutaneous tumours, but much lower percentages of cells that formed tumours after intravenous or intrasplenic transplantation, particularly among inefficiently metastasizing melanomas. Melanoma cells in the blood and visceral organs experienced oxidative stress not observed in established subcutaneous tumours. Successfully metastasizing melanomas underwent reversible metabolic changes during metastasis that increased their capacity to withstand oxidative stress, including increased dependence on NADPH-generating enzymes in the folate pathway. Antioxidants promoted distant metastasis in NSG mice. Folate pathway inhibition using low-dose methotrexate, ALDH1L2 knockdown, or MTHFD1 knockdown inhibited distant metastasis without significantly affecting the growth of subcutaneous tumours in the same mice. Oxidative stress thus limits distant metastasis by melanoma cells *in vivo*.**

Circulating cancer cells are commonly observed in the blood of patients and mice with various cancers<sup>1–4</sup>. However, metastasis is a very inefficient process<sup>5</sup> in which few disseminated cancer cells survive and even fewer proliferate<sup>6–8</sup>. Some patients can have circulating cancer cells in their blood without evidence of metastasis or worse outcomes<sup>9–11</sup>.

Epithelial cells undergo cell death when they detach from extracellular matrix in culture as a result of reduced glucose uptake, ATP depletion and oxidative stress<sup>12,13</sup>. Oncogenic signalling can promote their survival by increasing glucose uptake and flux through the pentose phosphate pathway, which generates NADPH and regenerates glutathione, a buffer against oxidative stress<sup>14</sup>. Glutathione is necessary for the initiation of some cancers, and antioxidants can promote cancer initiation and progression<sup>15–18</sup>. Cancer cells thus undergo genetic changes within primary tumours that increase their capacity to withstand oxidative stress, raising the question of whether additional adaptations are required during metastasis. Breast and lung cancer cell lines undergo metabolic changes during invasion in culture and metastasis *in vivo* that would be expected to reduce the generation of reactive oxygen species (ROS)<sup>18–23</sup>. Nonetheless, it is unknown whether ROS levels change in metastasizing cells *in vivo* or whether this limits distant metastasis. In fact, antioxidants inhibit the metastasis of some cancer cell lines, raising the possibility that ROS promotes metastasis in certain contexts<sup>24–26</sup>.

We addressed these issues by studying melanomas from several patients that were xenografted into NSG mice. Melanoma metastasis in this assay is predictive of clinical outcome in patients: stage III melanomas that metastasize efficiently in NSG mice go on to form distant metastases in patients, despite surgical resection, whereas melanomas that metastasize inefficiently in mice are usually cured by surgery in patients<sup>27</sup>.

## Blood and viscera are hostile to metastasis

We obtained four efficiently (UT10, M481, M405 and M514) and four inefficiently (M597, M528, M610 and M498) metastasizing

melanomas from patients. All expressed melanoma markers (Extended Data Fig. 1). The efficiently metastasizing melanomas formed distant metastases in patients and in NSG mice after subcutaneous injection (Extended Data Fig. 2a). Conversely, inefficiently metastasizing melanomas did not form distant metastases in patients or macrometastases in NSG mice (Extended Data Fig. 2a).

The efficiently and inefficiently metastasizing melanomas did not significantly differ with regard to the frequency of cells that formed tumours after subcutaneous injection in NSG mice (Table 1) or the rate at which these tumours grew (Extended Data Fig. 2b). One in eight cells from efficiently metastasizing and one in eleven cells from inefficiently metastasizing melanomas formed tumours after subcutaneous injection (Table 1). We often detected circulating melanoma cells by flow cytometry in the blood of mice with efficiently metastasizing, but not inefficiently metastasizing, melanomas (Extended Data Fig. 3a, b).

To test whether the main difference between efficiently and inefficiently metastasizing melanomas is the ability to enter the circulation, we intravenously injected 10, 100, 1,000 or 10,000 cells from efficiently and inefficiently metastasizing melanomas into NSG mice. The melanoma cells were marked by luciferase expression, allowing us to confirm micro- and macrometastases by bioluminescence imaging. Efficiently metastasizing melanomas from all four patients formed macrometastases in several visceral organs (Extended Data Fig. 3c). Limiting dilution analysis indicated that at least 1 in 235 cells formed tumours after intravenous injection (Table 1). By contrast, only 1 in 2,540 cells from inefficiently metastasizing melanomas formed tumours after intravenous injection (Table 1 and Extended Data Fig. 3d). Therefore, the ability to enter circulation is not the only factor that limits distant metastasis.

These data also demonstrate that even efficiently metastasizing melanomas more readily formed tumours after subcutaneous injection (1 in 8 cells) as compared to intravenous injection (1 in 235 cells; Table 1). This was true even when efficiently metastasizing melanoma

<sup>1</sup>Children's Research Institute and the Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>2</sup>Department of Surgery, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>3</sup>Department of Dermatology, University of Michigan, Ann Arbor, Michigan 48109-2216, USA. <sup>4</sup>Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.



**Table 1 | Fraction of cells that formed tumours after injection**

Melanoma	Subcutaneous	Intravenous	Intrasplenic
Inefficient	1/11	1/2,540	1/3,677
Efficient	1/8	1/235	1/173

The fraction of melanoma cells that formed tumours after subcutaneous, intravenous or intrasplenic injection. Melanoma cells more readily formed tumours after subcutaneous injection (for example, 1 in 8 cells formed tumours) as compared to intravenous (for example, 1 in 235 cells) or intrasplenic (for example, 1 in 173 cells) injection. Limit dilution analysis of data in Extended Data Fig. 3 to infer the minimum frequency of tumour-forming melanoma cells after subcutaneous, intravenous or intrasplenic transplantation into NSG mice. Each mean value reflects 5 (subcutaneous), 10 (intravenous), or 4 (intrasplenic) independent experiments using melanomas obtained from 8 different patients (see details in Extended Data Fig. 3). Statistical significance was assessed by a Chi-square test using the ELDA software<sup>46</sup> (\*\*\* $P < 0.0001$ ).

cells were injected subcutaneously without Matrigel (1 in 60 cells formed tumours;  $P < 0.001$ ). This suggests that the blood was more hostile to melanoma cells than the subcutaneous environment.

If distant metastasis is limited mainly by survival in the blood, direct injection into a visceral organ should increase tumour formation. To test this, we injected efficiently and inefficiently metastasizing melanomas into the spleens of NSG mice. Efficiently metastasizing melanomas from two out of three patients formed macrometastases in several visceral organs in most mice (Extended Data Fig. 3e and Table 1; at least 1 in 173 cells formed tumours after intrasplenic injection). By contrast, only 1 in 3,677 cells from inefficiently metastasizing melanomas formed tumours after intrasplenic injection (Extended Data Fig. 3f and Table 1). The ability to survive in the circulation is therefore not the only factor that limits distant metastasis. The spleen is also a relatively hostile environment for melanoma cells.

### Reversible tropic changes during metastasis

To test whether melanoma cells undergo changes in properties during metastasis, we obtained efficiently metastasizing melanoma cells from 12 donor mice that had been grafted with melanomas from 3 patients (M481, M405 and UT10). We compared the capacity of melanoma cells from subcutaneous tumours versus the blood versus metastatic liver nodules (2–5 mm in diameter) in the same donor mice to form tumours after subcutaneous, intravenous or intrasplenic injection in recipient mice (Table 2a; see Extended Data Fig. 3g for experimental

design). After subcutaneous injection, melanoma cells from subcutaneous tumours were significantly better at forming tumours (1 in 14 cells formed tumours) than melanoma cells from the blood (1 in 63 cells) or metastatic nodules (1 in 55 cells; Table 2a). By contrast, after intrasplenic injection, melanoma cells from metastatic nodules were significantly better at forming tumours (1 in 130 cells) than melanoma cells from the blood (1 in 372 cells) or subcutaneous tumours (1 in 708 cells; Table 2a). This suggests melanoma cells adapt to the sites of metastasis as they metastasize.

To test whether the changes were irreversible (for example, genetic) or reversible (for example, epigenetic or metabolic), we tested whether small numbers of metastatic melanoma cells would reacquire subcutaneous properties after being passaged subcutaneously for a short period of time. We subcutaneously transplanted 100 cells from subcutaneous tumours, the blood or metastatic liver nodules of the same donor mice into primary recipient mice and allowed them to form tumours for up to 12 weeks (Table 2b; all injections formed tumours). Then we retransplanted melanoma cells from the subcutaneous tumours into subcutaneous, intravenous or intrasplenic sites in secondary recipient mice (Extended Data Fig. 3h shows experimental design). Melanoma cells that originated in all sites formed subcutaneous tumours with high efficiency in secondary recipient mice after being passaged subcutaneously (Table 2b). Melanoma cells that originated in all sites formed metastatic tumours with low efficiency after intravenous or intrasplenic injections in secondary recipient mice after being passaged subcutaneously (Table 2b). The changes in tumorigenic tropism during metastasis are thus reversible. Since these reversible changes were observed very consistently, even in small numbers of melanoma cells, these changes cannot reflect selection for rare genetic events during metastasis followed by reversion mutations after subcutaneous re-transplantation.

### Oxidative stress limits distant metastasis

We performed liquid chromatography–tandem mass spectrometry (LC–MS/MS) metabolomics on subcutaneous tumours and visceral metastatic nodules from the same NSG mice transplanted with efficiently metastasizing melanomas derived from four patients. In two independent experiments, unsupervised clustering of metabolomics

**Table 2 | Melanoma cells undergo reversible changes in tropism during metastasis**

#### a Transplantation in primary recipient mice

Source of melanoma cells in donor mice	Transplantation site in primary recipient mice	Tumorigenic frequency
Subcutaneous tumour cells	Subcutaneous	1/14
Circulating cells		1/63
Metastatic nodules		1/55
Subcutaneous tumour cells	Intravenous	1/1,282
Circulating cells		1/4,270
Metastatic nodules		1/627
Subcutaneous tumour cells	Intrasplenic	1/708
Circulating cells		1/372
Metastatic nodules		1/130

#### b Transplantation in secondary recipient mice

Source of melanoma cells in donor mice	Passage site in primary recipient mice	Transplantation site in secondary recipient mice	Tumorigenic frequency
Subcutaneous tumour cells	Subcutaneous	Subcutaneous	1/11
Circulating cells			1/17
Metastatic nodules			1/11
Subcutaneous tumour cells	Subcutaneous	Intravenous	1/584
Circulating cells			1/566
Metastatic nodules			1/899
Subcutaneous tumour cells	Subcutaneous	Intrasplenic	1/843
Circulating cells			1/307
Metastatic nodules			1/584

a, Limiting dilution analysis of the fraction of melanoma cells from subcutaneous tumours, the blood (circulating cells), or metastatic nodules that formed tumours after subcutaneous, intravenous, or intrasplenic transplantation. These data reflect three independent experiments performed with efficiently metastasizing melanomas (M405, M481 and UT10;  $n = 10$  mice/melanoma/melanoma cell source/transplantation site for a total of 270 mice).

b, Limiting dilution analysis of the fraction of melanoma cells from subcutaneous tumours, the blood (circulating cells), or metastatic nodules that formed tumours after being passaged subcutaneously in primary recipient mice and then transplanted subcutaneously, intravenously, or intrasplenically into secondary recipient mice. These data reflect one experiment performed with efficiently metastasizing melanoma cells (M481;  $n = 8$ –10 mice/melanoma cell source/transplantation site for a total of 85 mice).

Statistical significance was assessed by a Chi-square test using ELDA software<sup>46</sup>. \* $P < 0.05$ ; \*\* $P < 0.005$ ; \*\*\* $P < 0.0005$ .

data showed that metastatic nodules obtained from the liver, pancreas and kidney almost always clustered together, distinct from subcutaneous tumours, irrespective of the xenograft line (Extended Data Fig. 4a, b; complete metabolomics data are shown in Extended Data Tables 1 and 2). Among subcutaneous tumours and among metastatic nodules, samples clustered by patient. We obtained similar results when melanoma cells were isolated by flow cytometry, excluding mouse cells (Extended Data Fig. 4c). The metabolic differences between subcutaneous tumours and visceral metastases appeared to be largely reversible as metabolomics analysis of tumours from Table 2b showed that subcutaneous tumours clustered together irrespective of whether they arose from the transplantation of subcutaneous, circulating or metastatic cells (Extended Data Fig. 4d, e).

In a total of six independent experiments that involved four different technical approaches, the glutathione (GSH) to oxidized glutathione (GSSG) ratio was always significantly higher in subcutaneous tumours than in metastatic nodules or circulating melanoma cells (Fig. 1a, b and Extended Data Fig. 4f). This was true irrespective of whether melanoma cells were isolated by dissection or by flow cytometry (to eliminate stromal cells). The lower GSH/GSSG ratio in circulating melanoma cells and metastatic nodules suggested that metastasizing cells experienced oxidative stress not observed in established subcutaneous tumours, and they consumed GSH in an effort to maintain redox homeostasis.

Consistent with this, cytoplasmic ROS levels were significantly higher in circulating melanoma cells and visceral metastatic nodules than in subcutaneous tumours (Fig. 1c). Mitochondrial ROS levels were significantly higher in visceral metastatic nodules than in circulating melanoma cells and subcutaneous tumours (Fig. 1d).

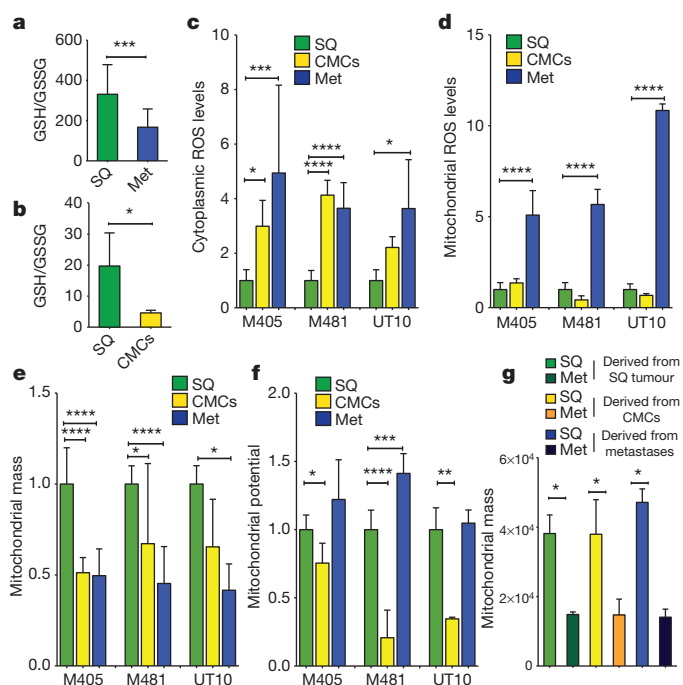
Mitochondrial respiration is one of the main sources of ROS. Mitochondrial mass declined significantly in circulating melanoma cells and metastatic nodules as compared to subcutaneous tumours (Fig. 1e). Mitochondrial membrane potential also declined significantly in circulating melanoma cells as compared to subcutaneous tumours but not in metastatic nodules (Fig. 1f). These data raised the possibility that mitochondrial function is reduced in circulating melanoma cells in an effort to reduce ROS generation.

These changes in redox status between subcutaneous and metastasizing cells appeared to be reversible. Even though the GSH/GSSG ratio was always higher in subcutaneous tumours than in metastatic nodules, the GSH/GSSG ratio in subcutaneous tumours derived from the transplantation of metastatic cells was even higher than the GSH/GSSG ratio in subcutaneous tumours derived from the transplantation of subcutaneous cells (Extended Data Fig. 4g).

Changes in mitochondrial mass between subcutaneous and metastatic melanoma cells were also reversible. Subcutaneous tumours always exhibited significantly higher mitochondrial mass than metastatic nodules, irrespective of whether the subcutaneous tumours arose from the transplantation of subcutaneous tumour cells, circulating cells or metastatic cells (Fig. 1g and Extended Data Fig. 4h).

To test whether oxidative stress limits melanoma metastasis, we subcutaneously transplanted efficiently metastasizing melanoma cells derived from three patients into NSG mice and treated the mice with daily subcutaneous injections of the antioxidant *N*-acetyl-cysteine (NAC; 200 mg kg<sup>-1</sup> day<sup>-1</sup>). In no case did NAC treatment significantly affect the growth of established subcutaneous tumours (Fig. 2a), but it significantly increased the frequency of melanoma cells in the blood of mice transplanted with M405 and UT10 (Fig. 2b), and significantly increased metastatic disease burden in mice with all three melanomas (Fig. 2c). Oxidative stress therefore limits the metastasis of melanoma cells *in vivo*.

Among efficiently metastasizing melanomas from three patients, NAC pre-treatment of cells and administration to mice increased tumour formation after intravenous injection by tenfold ( $P < 0.0001$ ; Extended Data Table 3a). Among inefficiently metastasizing melanomas from two patients, tumours only arose from cells

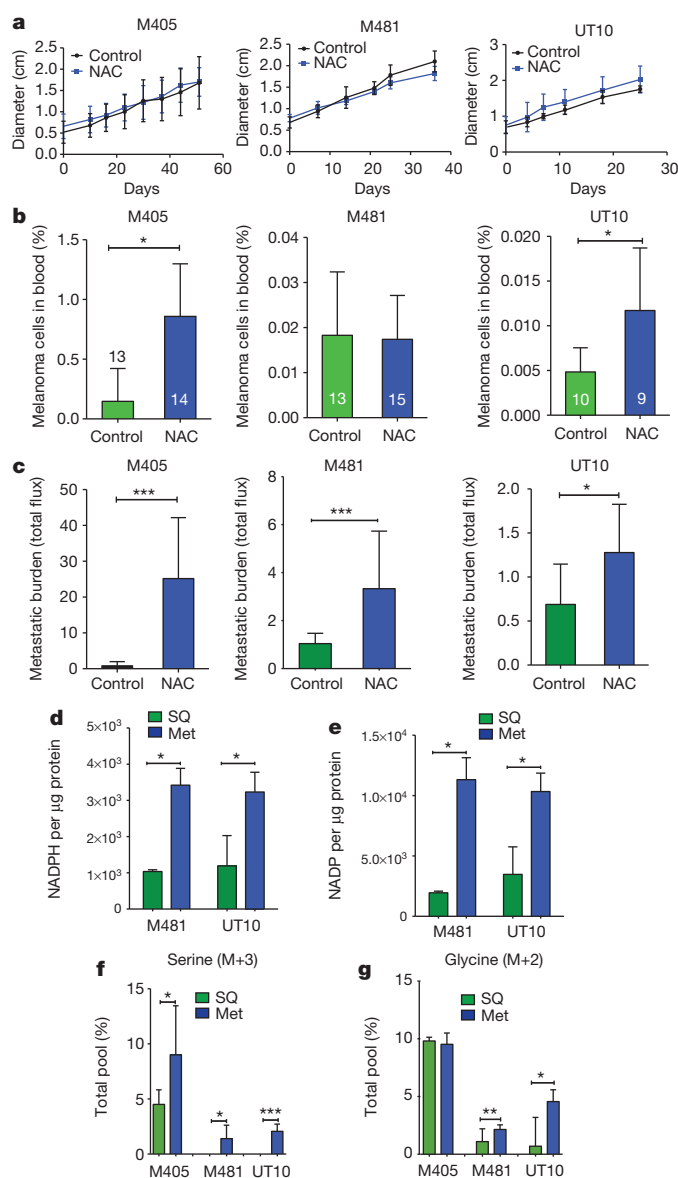


**Figure 1 | Metastasizing melanoma cells experience high levels of oxidative stress.** **a**, GSH/GSSG ratio in subcutaneous tumours (SQ) as compared to metastatic nodules (Met) ( $n = 15$  mice from two independent experiments with three melanomas (M481, M405 and UT10); note that extractions were performed with 0.1% formic acid to prevent spontaneous oxidation<sup>47</sup>). Total amounts of GSH and GSSG are shown in Extended Data Fig. 5h, i. **b**, GSH/GSSG ratio in subcutaneous tumours as compared to circulating melanoma cells ( $n = 7$  mice from three independent experiments with two melanomas (M405 and UT10); these were different experiments than those in **a**, performed under different technical conditions). **c**, **d**, Cytoplasmic (**c**) and mitochondrial (**d**) ROS levels in dissociated melanoma cells from subcutaneous tumours, the blood (circulating melanoma cells, CMCs), and metastatic nodules obtained from the same mice ( $n = 9$  mice from three independent experiments using three different melanomas). **e**, **f**, Mitochondrial mass (**e**) and mitochondrial membrane potential (**f**) in dissociated melanoma cells from subcutaneous tumours, the blood, and metastatic nodules obtained from the same mice ( $n = 6$  mice from two independent experiments using three different melanomas). **g**, Melanoma cells underwent reversible changes in mitochondrial mass during metastasis: mitochondrial mass in dissociated melanoma cells from subcutaneous tumours versus metastatic nodules obtained from the same mice transplanted with subcutaneous, circulating or metastatic melanoma cells. All data represent mean  $\pm$  s.d. Statistical significance was assessed using two-tailed Student's *t*-tests (**a** and **b**) and one-way analyses of variance (ANOVAs) followed by Dunnett's tests for multiple comparisons (**c**–**g**). \* $P < 0.05$ ; \*\*\* $P < 0.0005$ ; \*\*\*\* $P < 0.00005$ .

treated with NAC (Extended Data Table 3b). Oxidative stress thus limits tumorigenesis by circulating melanoma cells.

## Metabolic adaptations during metastasis

We proposed that successfully metastasizing cells undergo reversible metabolic changes that increase their capacity to withstand oxidative stress. One such adaptation that could promote survival would be increased GSH regeneration<sup>28–30</sup>. NADPH is needed to convert GSSG into GSH and increased production of NADPH promotes the regeneration of GSH<sup>28–30</sup>. In fact, we observed significantly higher levels of NADPH and NADP in metastatic cells than in subcutaneous tumours (Fig. 2d, e). The NADPH/NADP ratio was significantly reduced in metastases as compared to subcutaneous tumours for one melanoma (M481) but not for a second (UT10; Extended Data Fig. 5e). The higher levels of NADPH in metastases raised the possibility that metastasizing cells generate more NADPH to increase their capacity to regenerate GSH. The oxidative stress in these cells would be predicted to consume more NADPH, potentially explaining



**Figure 2 | Melanoma cell metastasis, but not subcutaneous tumour growth, is promoted by antioxidants *in vivo*.** **a–c**, Growth of established subcutaneous tumours in NSG mice treated with either PBS (control) or NAC by daily subcutaneous injection. Tumour diameter source data are shown in Supplementary Fig. 1. Frequency of circulating melanoma cells in the blood (**b**) and metastatic disease burden (**c**) assessed based on total bioluminescence signal from the visceral organs of the same mice. Data in **a–c** represent eight independent experiments, with total replicates/treatment shown in the bars of **b**. A single representative experiment per melanoma is shown in **a** owing to the difficulty of reflecting tumour growth measurements from independent experiments in the same graph. No statistically significant differences among treatments were observed in subcutaneous tumour growth in any experiment. **d, e**, Levels of NADPH (**d**) and NADP (**e**) in subcutaneous tumours versus metastatic nodules ( $n = 4$  mice from two independent experiments with M481 and UT10). **f, g**, *In vivo* isotope tracing of uniformly  $^{13}\text{C}$ -labelled glucose into serine (**f**) and glycine (**g**) in subcutaneous tumours versus metastatic nodules from the same mice ( $n = 6$  mice in two independent experiments for M405;  $n = 3$  mice in one experiment for each of M481 and UT10). The fragments for uniformly labelled serine (M + 3) and glycine (M + 2), which come from labelled glucose via *de novo* serine synthesis, are shown. All data represent mean  $\pm$  s.d. Statistical significance was assessed using two-tailed Student's *t*-tests (**d–f, h** and **i**), or the Mann–Whitney test (**g**, due to unequal variance), and repeated-measures two-way ANOVAs (**a–c**). \* $P < 0.05$ ; \*\* $P < 0.005$ ; \*\*\* $P < 0.0005$ ; \*\*\*\* $P < 0.00005$ .

why the NADPH/NADP ratio did not necessarily change despite increases in the amounts of NADPH and NADP.

Metastatic nodules had high levels of purine intermediates relative to subcutaneous tumours in the same mice (Extended Data Tables 1 and 2). This could reflect changed folate pathway activity. Moreover, folate metabolism is a major source of NADPH for oxidative stress management<sup>29–31</sup>, raising the question of whether the folate pathway promotes distant metastasis. NADPH regeneration by the folate pathway involves the conversion of serine to glycine and the donation of a methyl group to tetrahydrofolate (see Extended Data Fig. 5j). Serine can be imported into cells as well as produced *de novo* from glucose via the glycolytic intermediate 3-phosphoglycerate (3-PG). Increased *de novo* serine synthesis promotes the growth of some melanomas and breast cancers<sup>32,33</sup>. To test for alterations of these pathways during metastasis, we administered uniformly  $^{13}\text{C}$ -labelled glucose to NSG mice bearing melanomas derived from three patients. We observed significantly increased  $^{13}\text{C}$ -labelling of serine and glycine in metastatic tumours as compared to subcutaneous tumours (Fig. 3f, g and Extended Data Fig. 5c, d). However, we observed no differences between subcutaneous tumours and metastases in terms of the fractional enrichments of uniformly labelled lactate or 3-PG, the precursor used for *de novo* serine synthesis (Extended Data Fig. 5a, b). Metastatic tumours thus display enhanced contribution of glucose carbon to tissue serine and glycine levels, suggesting increased flux through the folate pathway.

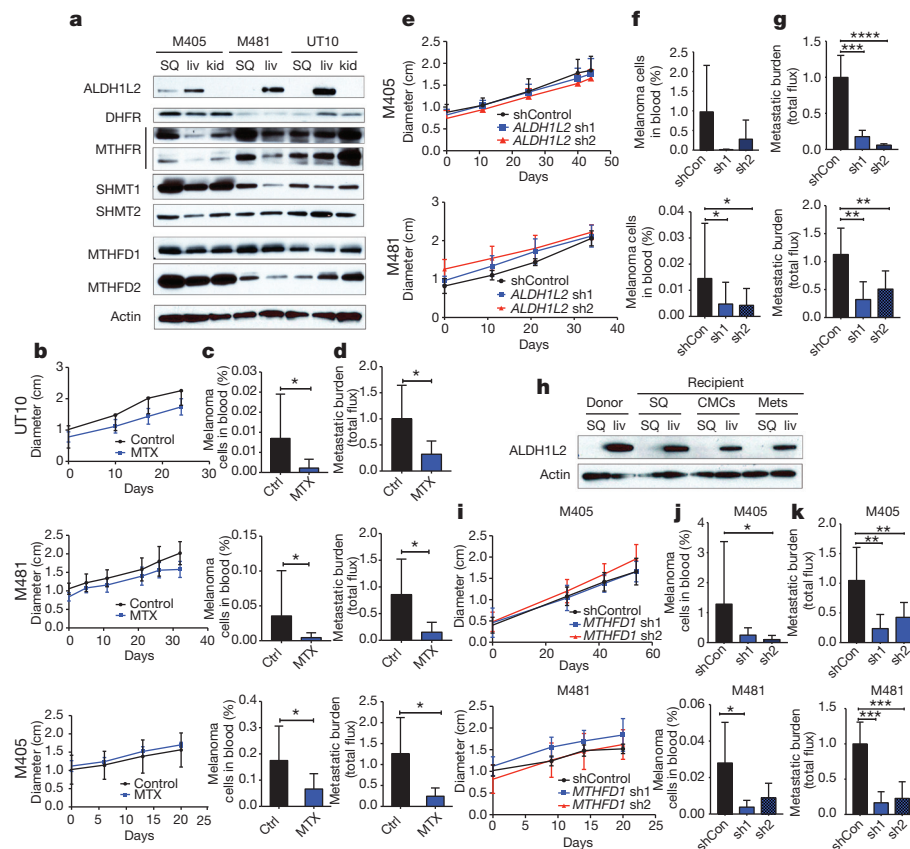
Western blotting and immunofluorescence analysis of key enzymes in the folate pathway showed a striking increase in the expression of the NADPH-regenerating enzyme ALDH1L2 in liver, pancreas and lung metastases compared to subcutaneous tumours, although not in kidney metastases (Fig. 3a and Extended Data Fig. 6). Some metastases also showed reduced expression of an NADPH-consuming folate metabolism enzyme, MTHFR (Fig. 3a). The expression of other folate metabolism enzymes did not consistently change during metastasis, nor did a variety of other enzymes in antioxidant pathways (Fig. 3a and Extended Data Table 4).

We compared ALDH1L2 levels in subcutaneous tumours versus metastatic nodules from mice transplanted with subcutaneous, circulating or metastatic melanoma cells. ALDH1L2 was much more highly expressed in liver nodules than in subcutaneous tumours in both donor and recipient mice, irrespective of whether the subcutaneous tumours arose from subcutaneous, circulating or metastatic melanoma cells (Fig. 3h). The changes in ALDH1L2 expression during metastasis were thus reversible.

To test whether metastasizing cells are more dependent than subcutaneous cells on the folate pathway, we transplanted melanoma cells from three patients subcutaneously in NSG mice and treated with a low dose of methotrexate ( $1.25 \text{ mg kg}^{-1} \text{ day}^{-1}$ ), an inhibitor of dihydrofolate reductase (Extended Data Fig. 5j). We simultaneously provided thymidine ( $3 \text{ mg kg}^{-1} \text{ day}^{-1}$ ) and hypoxanthine ( $750 \mu\text{g mg}^{-1} \text{ day}^{-1}$ ) to the mice to ameliorate the effects of folate pathway inhibition on nucleotide metabolism. Methotrexate treatment in these conditions had no significant effect on the growth of subcutaneous tumours (Fig. 3b); however, the frequency of circulating melanoma cells in the blood of the same mice was significantly reduced (Fig. 3c), as was metastatic disease burden (Fig. 3d), in mice bearing all three melanomas. Metastasizing melanoma cells are therefore particularly sensitive to folate pathway inhibition.

Depletion of ALDH1L2 can decrease the GSH/GSSG ratio *in vitro*<sup>30</sup>. To test whether ALDH1L2 is required for metastasis, we identified two short hairpin RNAs (shRNAs) that knocked down ALDH1L2 expression in melanoma cells (Extended Data Fig. 5f). We infected melanoma cells derived from three patients with either of these shRNAs or a scrambled control shRNA, then injected the cells subcutaneously in NSG mice. Neither of the shRNAs against ALDH1L2 significantly affected the growth of subcutaneous tumours (Fig. 3e), but both shRNAs significantly reduced the frequency of circulating melanoma





**Figure 3 | During metastasis, some melanoma cells reversibly increase their expression of folate pathway enzymes that generate NADPH and folate pathway inhibition selectively impairs metastasis.** **a**, Western blot analysis of folate pathway enzymes in subcutaneous tumours versus metastatic liver (liv) and kidney (kid) nodules from NSG mice transplanted with three different melanomas. **b–d**, Growth of subcutaneous tumours in mice bearing three different melanomas (M405, M481 and UT10) treated with dimethylsulfoxide (DMSO; control (ctrl)) or methotrexate (MTX) ( $n = 5$  mice per treatment). The frequency of circulating melanoma cells in the blood (**c**) and metastatic disease burden (**d**) in the same mice ( $n = 10$  mice per treatment for each melanoma except  $n = 8$  for M405). Data in **b–d** reflect six independent experiments, but only one representative experiment per melanoma is shown in **b**. **e–g**, Growth of subcutaneous tumours in mice transplanted with two different melanomas expressing scrambled control shRNA versus two shRNAs against *ALDH1L2*. The frequency of circulating melanoma cells (**f**) and metastatic disease burden in visceral organs based on total bioluminescence signal (**g**). The data in **e–g** reflect six independent experiments ( $n = 10$  mice per shRNA for M405 and  $n = 19$  mice per shRNA for M481), but only one representative experiment per

melanoma is shown in **e**. **h**, Western blot analysis of *ALDH1L2* expression in subcutaneous tumours versus metastatic liver nodules from a donor mouse or from recipient mice subcutaneously transplanted with subcutaneous, circulating or metastatic melanoma cells from the donor mouse. The increase in *ALDH1L2* expression in metastatic liver nodules was reversible after subcutaneous transplantation. Data in **a** and **h** are from two independent experiments. **i–k**, Growth of subcutaneous tumours in mice transplanted with cells from two melanomas expressing either scrambled control shRNAs or two shRNAs against *MTHFD1*. Frequency of circulating melanoma cells (**j**) and metastatic disease burden in visceral organs (**k**) from the same mice. Data in **i–k** reflect four independent experiments, with a total of nine mice per control shRNA and ten mice per shRNA against *MTHFD1* for each melanoma. One representative experiment per melanoma is shown in **i**. All error bars represent s.d. Statistical significance was assessed using ANOVA followed by Dunnett's test for multiple comparisons (**f**, **g**, **j** and **k**), two-tailed Student's *t*-tests (**c** and **d**) and repeated measures two-way ANOVAs (**b**, **e** and **i**). \* $P < 0.05$ ; \*\* $P < 0.005$ ; \*\*\* $P < 0.0005$ ; \*\*\*\* $P < 0.00005$ . Tumour diameter and western blot source data are in Supplementary Fig. 1.

cells in the blood of mice bearing one of two melanomas (Fig. 3f) and overall metastatic disease burden in both melanomas (Fig. 3g). *ALDH1L2* thus promotes melanoma metastasis *in vivo*.

We performed a similar experiment on another NADPH-regenerating enzyme in the folate pathway, *MTHFD1* (Extended Data Fig. 5j). We identified two shRNAs that knocked down *MTHFD1* expression in melanoma cells (Extended Data Fig. 5g). Neither shRNA against *MTHFD1* significantly affected the growth of subcutaneous tumours (Fig. 3i), but they both reduced the frequencies of circulating melanoma cells in the blood (Fig. 3j) and metastatic disease burden in mice bearing both melanomas (Fig. 3k). *MTHFD1* thus promotes melanoma metastasis *in vivo*.

Our results suggest that few circulating cancer cells survive and proliferate<sup>6–8</sup> because of oxidative stress. Our results further suggest that metabolic pathways that regenerate NADPH and buffer oxidative stress represent therapeutic targets to impede distant metastasis. Metastasizing melanoma cells are more dependent on certain NADPH-regenerating folate pathway enzymes than subcutaneous

melanoma cells. However, these enzymes cannot account for the increase in NADP in metastatic nodules as compared to subcutaneous tumours (Fig. 2d, e). Therefore, metastasizing cells probably also undergo additional metabolic changes that promote the generation of NADP and NADPH.

Changes in the expression of individual folate pathway enzymes are unlikely to be sufficient to drive metastasis. Since the key changes driving distant metastasis were reversible adaptations rather than stochastic genetic changes, there is no reason why these changes would need to be driven by a single sufficient driver. Adaptive changes might involve coordinated changes in the expression of several enzymes to increase flux through metabolic pathways. Other metabolic pathways also contribute to the synthesis and regeneration of NADPH, including the pentose phosphate<sup>12,13</sup> and malate<sup>34</sup> pathways. These pathways could potentially also contribute to the ability of melanoma cells to survive during metastasis.

Our experiments showed that oxidative stress increases in metastasizing cells as compared to established subcutaneous tumours and

limits distant metastasis. However, we did not test whether oxidative stress also limits the initiation and early growth of primary cutaneous melanomas. Thus, transient oxidative stress could occur during the formation of primary tumours<sup>15–18</sup> in addition to limiting distant metastasis at a later stage of cancer progression.

ROS can cause oncogenic mutations and activate oncogenic pathways, raising the possibility that treatment with antioxidants could suppress the initiation or progression of some cancers<sup>28,35</sup>. Antioxidants or antioxidant enzymes suppress cancer initiation in some contexts<sup>36–38</sup> while increasing cancer initiation in other contexts<sup>15–17,39</sup>. Increasing dietary antioxidants has generally not reduced cancer incidence in clinical trials<sup>40</sup>. Dietary supplementation with antioxidants actually increased incidence and death from lung and prostate cancer<sup>41–43</sup>. Dietary supplementation with folate promotes the development and progression of breast cancer<sup>44,45</sup>. Our results suggest that antioxidants promote disease progression, at least in melanoma, by promoting metastasis.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 25 December 2014; accepted 17 September 2015.**

**Published online 14 October 2015.**

- Yu, M. *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**, 580–584 (2013).
- Stott, S. L. *et al.* Isolation and characterization of circulating tumor cells from patients with localized and metastatic prostate cancer. *Sci. Transl. Med.* **2**, 25ra23 (2010).
- Yu, M. *et al.* Cancer therapy. *Ex vivo* culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* **345**, 216–220 (2014).
- Nagrath, S. *et al.* Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* **450**, 1235–1239 (2007).
- Vanharanta, S. & Massague, J. Origins of metastatic traits. *Cancer Cell* **24**, 410–421 (2013).
- Luzzi, K. J. *et al.* Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am. J. Pathol.* **153**, 865–873 (1998).
- Cameron, M. D. *et al.* Temporal progression of metastasis in lung: cell survival, dormancy, and location dependence of metastatic inefficiency. *Cancer Res.* **60**, 2541–2546 (2000).
- Kienast, Y. *et al.* Real-time imaging reveals the single steps of brain metastasis formation. *Nature Med.* **16**, 116–122 (2010).
- Engell, H. C. Cancer cells in the blood; a five to nine year follow up study. *Ann. Surg.* **149**, 457–461 (1959).
- Griffiths, J. D., McKinna, J. A., Rowbotham, H. D., Tsolakidis, P. & Salsbury, A. J. Carcinoma of the colon and rectum: circulating malignant cells and five-year survival. *Cancer* **31**, 226–236 (1973).
- Salsbury, A. J. The significance of the circulating cancer cell. *Cancer Treat. Rev.* **2**, 55–72 (1975).
- Debnath, J. & Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nature Rev. Cancer* **5**, 675–688 (2005).
- Debnath, J. *et al.* The role of apoptosis in creating and maintaining luminal space within normal and oncogene-expressing mammary acini. *Cell* **111**, 29–40 (2002).
- Schafer, Z. T. *et al.* Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. *Nature* **461**, 109–113 (2009).
- Harris, I. S. *et al.* Glutathione and thioredoxin antioxidant pathways synergize to drive cancer initiation and progression. *Cancer Cell* **27**, 211–222 (2015).
- Sayin, V. I. *et al.* Antioxidants accelerate lung cancer progression in mice. *Sci. Transl. Med.* **6**, 221ra215 (2014).
- DeNicola, G. M. *et al.* Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* **475**, 106–109 (2011).
- Dey, S. *et al.* ATF4-dependent induction of heme oxygenase 1 prevents anoikis and promotes metastasis. *J. Clin. Invest.* **125**, 2592–2608 (2015).
- Dong, C. *et al.* Loss of FBP1 by Snail-mediated repression provides metabolic advantages in basal-like breast cancer. *Cancer Cell* **23**, 316–331 (2013).
- Kamarajugadda, S. *et al.* Manganese superoxide dismutase promotes anoikis resistance and tumor metastasis. *Cell Death Dis.* **4**, e504 (2013).
- Qu, Y. *et al.* Thioredoxin-like 2 regulates human cancer cell growth and metastasis via redox homeostasis and NF- $\kappa$ B signaling. *J. Clin. Invest.* **121**, 212–225 (2011).
- Chen, E. I. *et al.* Adaptation of energy metabolism in breast cancer brain metastases. *Cancer Res.* **67**, 1472–1486 (2007).
- Lu, X., Bennet, B., Mu, E., Rabinowitz, J. & Kang, Y. Metabolomic changes accompanying transformation and acquisition of metastatic potential in a syngeneic mouse mammary tumor model. *J. Biol. Chem.* **285**, 9317–9321 (2010).
- Porporato, P. E. *et al.* A mitochondrial switch promotes tumor metastasis. *Cell Rep.* **8**, 754–766 (2014).
- LeBleu, V. S. *et al.* PGC-1 $\alpha$  mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. *Nature Cell Biol.* **16**, 992–1003 (2014).
- Ishikawa, K. *et al.* ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis. *Science* **320**, 661–664 (2008).
- Quintana, E. *et al.* Human melanoma metastasis in NSG mice correlates with clinical outcome in patients. *Sci. Transl. Med.* **4**, 159ra149 (2012).
- Gorrini, C., Harris, I. S. & Mak, T. W. Modulation of oxidative stress as an anticancer strategy. *Nature Rev. Drug Discov.* **12**, 931–947 (2013).
- Lewis, C. A. *et al.* Tracing compartmentalized NADPH metabolism in the cytosol and mitochondria of mammalian cells. *Mol. Cell* **55**, 253–263 (2014).
- Fan, J. *et al.* Quantitative flux analysis reveals folate-dependent NADPH production. *Nature* **510**, 298–302 (2014).
- Ye, J. *et al.* Serine catabolism regulates mitochondrial redox control during hypoxia. *Cancer Discovery* **4**, 1406–1417 (2014).
- Locasale, J. W. *et al.* Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nature Genet.* **43**, 869–874 (2011).
- Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–350 (2011).
- Jiang, P., Du, W., Mancuso, A., Wellen, K. E. & Yang, X. Reciprocal regulation of p53 and malic enzymes modulates metabolism and senescence. *Nature* **493**, 689–693 (2013).
- Chandel, N. S. & Tuveson, D. A. The promise and perils of antioxidants for cancer patients. *N. Engl. J. Med.* **371**, 177–178 (2014).
- Gao, P. *et al.* HIF-dependent antitumorigenic effect of antioxidants *in vivo*. *Cancer Cell* **12**, 230–238 (2007).
- Teoh-Fitzgerald, M. L., Fitzgerald, M. P., Zhong, W., Askeland, R. W. & Domann, F. E. Epigenetic reprogramming governs EcSOD expression during human mammary epithelial cell differentiation, tumorigenesis and metastasis. *Oncogene* **33**, 358–368 (2014).
- Glaser, A. & Chandel, N. S. Targeting antioxidants for cancer therapy. *Biochem. Pharmacol.* **92**, 90–101 (2014).
- Glaser, A., Sena, L. A., Diebold, L. P., Mazar, A. P. & Chandel, N. S. Targeting SOD1 reduces experimental non-small-cell lung cancer. *J. Clin. Invest.* **124**, 117–128 (2014).
- Fortmann, S. P. *et al.* Vitamin, mineral, and multivitamin supplements for the primary prevention of cardiovascular disease and cancer: a systematic evidence review for the U.S. preventive services task force U.S. preventive services task force. *Ann Intern Med.* **159**, 824–834 (2013).
- The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N. Engl. J. Med.* **330**, 1029–1035 (1994).
- Klein, E. A. *et al.* Vitamin E and the risk of prostate cancer: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *J. Am. Med. Assoc.* **306**, 1549–1556 (2011).
- Goodman, G. E. *et al.* The Beta-Carotene and Retinol Efficacy Trial: incidence of lung cancer and cardiovascular disease mortality during 6-year follow-up after stopping beta-carotene and retinol supplements. *J. Natl. Cancer Inst.* **96**, 1743–1750 (2004).
- Dehghan Maneshadi, S. *et al.* Folic acid supplementation promotes mammary tumor progression in a rat model. *PLoS ONE* **9**, e84635 (2014).
- Ebbing, M. *et al.* Cancer incidence and mortality after treatment with folic acid and vitamin B12. *J. Am. Med. Assoc.* **302**, 2119–2126 (2009).
- Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** S.J.M. is a Howard Hughes Medical Institute (HHMI) Investigator, the Mary McDermott Cook Chair in Pediatric Genetics, the director of the Hamon Laboratory for Stem Cells and Cancer, and a Cancer Prevention and Research Institute of Texas Scholar. We thank K. Correll and M. Gross for mouse colony management; N. Loof and the Moody Foundation Flow Cytometry Facility. We thank N. Meireles and the University of Michigan Melanoma Biobank, for Biobank database and melanoma clinical data management.

**Author Contributions** E.P. and S.J.M. conceived the project, designed, and interpreted experiments. M.A. and R.J.D. participated in the design and interpretation of experiments related to metabolic mechanisms. E.P. performed all the experiments with technical assistance from S.E.H., metabolomics assistance from M.A. and imaging assistance from M.M.M. Z.H. and M.A. developed metabolomics methods. Z.Z. assisted with statistical analyses. A.M.L. and T.M.J. provided melanoma specimens and associated clinical data. E.P., M.A. and S.J.M. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.M. ([sean.morrison@utsouthwestern.edu](mailto:sean.morrison@utsouthwestern.edu)).

## METHODS

**Obtaining melanomas and enzymatic dissociation.** Melanoma specimens were obtained with informed consent from all patients according to protocols approved by the Institutional Review Boards of the University of Michigan Medical School (IRB MED approvals HUM00050754 and HUM00050085; see ref. 27) and the University of Texas Southwestern Medical Center. Tumours were dissociated in Sterile Closed System Tissue Grinders (SKS Science) in enzymatic digestion medium containing 200 U ml<sup>-1</sup> collagenase IV (Worthington) for 20 min at 37 °C. DNase (50–100 U ml<sup>-1</sup>) was added to reduce clumping of cells during digestion. Cells were filtered with a 40-µm cell strainer to obtain a single-cell suspensions.

**Cell labelling and sorting.** All melanomas used in this study stably expressed DsRed and luciferase so that the melanoma cells could be unambiguously distinguished from mouse cells by flow cytometry and by bioluminescence imaging. When isolated by flow cytometry, cells were also stained with antibodies against mouse CD45 (30-F11-APC, eBiosciences), mouse CD31 (390-APC, Biolegend), Ter119 (TER-119-APC, eBiosciences) and human HLA-A, -B, -C (G46-2.6-FITC, BD Biosciences) to select live human melanoma cells and to exclude contaminating mouse endothelial and haematopoietic cells. Live human melanoma cells were thus isolated by flow cytometry by sorting cells that were positive for DsRed and HLA and negative for mouse CD45, Ter119 and CD31. All antibody labelling was performed for 20 min on ice, followed by washing and centrifugation. Before flow cytometric analysis, cells were re-suspended in staining medium (L15 medium containing bovine serum albumin (1 mg ml<sup>-1</sup>), 1% penicillin/streptomycin, and 10 mM HEPES, pH 7.4) containing 4',6-diamidino-2-phenylindole (DAPI; 5 µg ml<sup>-1</sup>; Sigma) to eliminate dead cells from sorts and analyses. Sorts and analyses were performed using a FACS Aria flow cytometer (Becton Dickinson). After sorting, an aliquot of sorted melanoma cells was always reanalysed to check for purity, which was usually greater than 95%. For analysis of circulating melanoma cells, blood was collected from each mouse by cardiac puncture with a syringe pretreated with citrate-dextrose solution (Sigma). Red blood cells were precipitated by Ficoll sedimentation according to the manufacturer's instructions (Ficoll Paque Plus, GE Healthcare). Remaining cells were washed with Hanks' balanced salt solution (Invitrogen) before antibody staining and flow cytometric analysis. For limiting dilution analysis, cells for each mouse were sorted into individual wells of 96-well V-bottomed plates containing staining medium and loaded into syringes directly from the well (one well into one syringe into one mouse).

**Transplantation of melanoma cells.** After sorting, cells were counted and resuspended in staining medium with 25% high-protein Matrigel (product 354248; BD Biosciences). Subcutaneous injections were performed into the right flank of NOD.CB17-Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl</sup>/SzJ (NSG) mice (Jackson Laboratory) in a final volume of 50 µl. Each mouse was transplanted with 100 melanoma cells unless otherwise specified. Tumour formation was evaluated regularly by palpation of the injection site, and the subcutaneous tumours were measured every 10 days until any tumour in the mouse cohort reached 2.5 cm in its largest diameter. Mice were monitored daily for signs of distress and euthanized when they exhibited distress according to a standard body condition score or within 24 h of their tumours reaching 2.5 cm in largest diameter, whichever came first. We adhered to this limit in all experiments. Organs were analysed visually and by bioluminescence imaging (see details below) for presence of macrometastases and micrometastases. These experiments were performed according to protocols approved by the Institutional Animal Care and Use Committee at the University of Texas Southwestern Medical Center (protocol 2011-0118). Intravenous injections were done by injecting cells into the tail vein of NSG mice in 100 µl of staining medium. For intrasplenic injections the mice were anaesthetized with isoflourane, then the left flank was shaved and disinfected with an ethanol wipe and iodine swab. An incision was made into the intraperitoneal cavity. The spleen was exposed with forceps and cells were injected slowly in a 40 µl volume of staining medium. The peritoneum was then sutured and skin was closed with clips. Mice were injected with buprenex before surgery and then again 12 h after surgery.

**Lentiviral transduction of human melanoma cells.** A bicistronic lentiviral construct carrying dsRed2 and luciferase (dsRed2-P2A-Luc) was generated (for bioluminescence imaging) and cloned into the FUW lentiviral expression construct. The primers that were used for generating this construct were: dsRed2 forward, 5'-CGACTCTAGAGGATCCatgtagactgagacgacgc-3' (capital letters indicate homology to FUW backbone); dsRed2 reverse, 5'-TCCACGCTCTCCAGCCTGCTTCAGCAGGCTGAAGTTAGTACTCCGCTTCCtggaacagtggtggc-3' (capital letters indicate P2A sequences); luciferase forward, 5'-GCCTGCTGAAGCAGGCTGGAGACGTGGAGAGAACCTGGACCTGGATCCatggaagacgcaaaaacataag-3' (capital letters indicate P2A sequences) and luciferase reverse, 5'-GCTTGATATCGAATTCttacacgcgcatcttccgc-3' (capital letters

indicate homology to FUW backbone). All constructs were generated using the In-Fusion HD cloning system (Clontech) and sequence verified.

For virus production, 0.9 µg of the appropriate plasmid and 1 µg of helper plasmids (0.4 µg pMD2G and 0.6 µg of psPAX2) were transfected into 293T cells using polyjet (Signagen) according to the manufacturer's instructions. Replication incompetent viral supernatants were collected 48 h after transfection and filtered through a 0.45-µm filter. Approximately 300,000 freshly dissociated melanoma cells were infected with viral supernatants supplemented with 10 µg ml<sup>-1</sup> polybrene (Sigma) for 4 h. Cells were then washed twice with staining medium, and about 25,000 cells (a mixture of infected and non-infected cells) were suspended in staining medium with 25% high-protein Matrigel (product 354248; BD Biosciences) then injected subcutaneously into NSG mice. After growing to 1–2 cm in diameter, tumours were excised and dissociated into single-cell suspensions, and luciferase-dsRed<sup>+</sup> or green fluorescent protein (GFP)<sup>+</sup> cells were collected by flow cytometry for injection into secondary recipients. Metastasis was monitored by bioluminescence imaging in secondary recipients. All shRNAs were expressed from a pGIPZ miRNA-based construct with TurboGFP from GE Dharmacon. For ALDH1L2, the following GE Dharmacon shRNA clones were used: V2LHS\_30207, V2LHS\_30209. For MTHFD1 the following GE Dharmacon shRNA clones were used: V2LHS\_216208 and V2LHS\_196832.

**Bioluminescence imaging.** Mice were injected with 100 luciferase-dsRed<sup>+</sup> cells on the right flank and monitored until tumour diameters approached 2.5 cm, at which point they were imaged along with an uninjected control mouse using an IVIS Imaging System 200 Series (Caliper Life Sciences) with Living Image software. Mice were injected intraperitoneally with 100 µl of PBS containing D-luciferin monopotassium salt (40 µg ml<sup>-1</sup>) (Biosynth) 5 min before imaging, followed by general anaesthesia 2 min before imaging. After imaging of the whole mouse, the mice were euthanized and individual organs were surgically removed and quickly imaged. The exposure time of images ranged from 10 to 60 s depending on signal intensity. The bioluminescence signal was quantified with 'region of interest' measurement tools in Living Image (Perkin Elmer) software. After imaging, tumours and organs were fixed in 10% neutral-buffered formalin for histopathology. For live imaging, mice were imaged once a month, and whole body bioluminescence was quantified using Living Image Software (Perkin Elmer).

**LC-MS/MS metabolomic analysis.** Mice were euthanized by cervical dislocation. Subcutaneous tumours and metastatic nodules were dissected, immediately homogenized in 80% methanol chilled with dry ice (Honeywell), vortexed vigorously, and metabolites were extracted overnight at –80 °C. The following day, samples were centrifuged at 13,000g for 15 min at 4 °C, the supernatant was collected, and metabolites from the pellet were re-extracted with 80% methanol at –80 °C for 4 h. After centrifugation, both supernatants were pooled and lyophilized using a SpeedVac (Thermo). To inhibit spontaneous oxidation, samples were extracted with 80% methanol containing 0.1% formic acid in some experiments<sup>47</sup>. Dried metabolites were reconstituted in 0.03% formic acid in water, vortexed and centrifuged, then the supernatant was analysed using liquid chromatography-tandem mass spectrometry (LC-MS/MS). A Nexera Ultra High Performance Liquid Chromatograph (UHPLC) system (Shimadzu) was used for liquid chromatography, with a Polar-RP HPLC column (150 × 2 mm, 4 µm, 80 Å, Phenomenex) and the following gradient: 0–3 min 100% mobile phase A; 3–15 min 100–0% A; 15–17 min 0% A; 17–18 min 0–100% A; 18–23 min 100% A. Mobile phase A was 0.03% formic acid in water. Mobile phase B was 0.03% formic acid in acetonitrile. The flow rate was 0.5 ml min<sup>-1</sup> and the column temperature was 35 °C. A triple quadrupole mass spectrometer (AB Sciex QTRAP 5500) was used for metabolite detection as previously described<sup>48</sup>. Chromatogram peak areas were integrated using Multiquant (AB Sciex). To measure GSH and GSSG levels, some metabolite extractions were performed with 0.1% formic acid in 80% methanol, to inhibit spontaneous GSH oxidation. To calculate GSH and GSSG amounts, a standard curve was prepared by adding known quantities of GSH and GSSG to tumour metabolite extracts.

**Isotope tracing with uniformly <sup>13</sup>C-labelled glucose.** Mice were injected intraperitoneally with 2 g kg<sup>-1</sup> body mass of uniformly <sup>13</sup>C-labelled glucose (Cambridge Isotopes) and were analysed 15, 30 and 60 min later. Mice were fasted for 14 h before the injection. In most experiments, subcutaneous tumours and metastatic nodules were surgically excised and homogenized in ice cold 50% methanol for GC-MS and in 80% dry ice-cold methanol for LC-MS analysis. Metabolites were extracted with three freeze-thaw cycles in liquid nitrogen. Supernatant was collected after a 15 min centrifugation at 13,000g at 4 °C and lyophilized. Metabolites were derivatized with trimethylsilyl (TMS) at 42 °C for 30 min for GC-MS analysis. <sup>13</sup>C-enrichment analysis was performed by GC-MS as previously described<sup>48</sup>. For LC-MS analysis, lyophilized samples were resuspended in either 0.03% formic acid in water or in 5 mM ammonium acetate in water depending on the method of analysis. For <sup>13</sup>C-enrichment analysis of



lactate, serine and glycine by LC-MS/MS, we used the liquid chromatography procedure described above for LC-MS/MS metabolomics analysis with the following modifications: the liquid chromatography gradient was 0–3 min 100% mobile phase A; 3–15 min 100–0% A; 15–17 min 0% A; 17–17.5 min 0–100% A; 17.5–20 min 100% A. For analysis of 3-PG, the liquid chromatography conditions were: mobile phase A, 5 mM ammonium acetate in water and mobile phase B, 5 mM ammonium acetate in acetonitrile, and a Fusion-RP HPLC column (150 × 2 mm, 4 µm, 80 Å, Phenomenex). The liquid chromatography gradient was: 0–3 min 100% mobile phase A; 3–9 min 100–0% A; 9–11 min 0% A; 11–12 min 0–100% A; 12–15 min 100% A. For metabolite detection a triple quadrupole mass spectrometer (AB Sciex QTRAP 5500) was used on multiple reaction monitoring mode as previously described, with some modifications<sup>48</sup>. The following transitions were used: positive mode: serine 106.1/60 (M + 1: 107.1/60 and 107.1/61, M + 2: 108.1/61 and 108.1/62, M + 3: 109.1/62), glycine 76/30 (M + 1 77/30 and 77/31, M + 2 78/31); negative mode: lactate 89/43 (M + 1 90/43 and 90/44, M + 2 91/44 and 91/45, M + 3 92/45), 3-PG 185/79 (M + 1 186/79, M + 2 187/79, M + 3 188/79) and 185/97 (M + 1 186/97, M + 2 187/97, M + 3 188/97). Unlabelled tissue was used as a negative control to confirm isotopic labelling in specific transitions.

**Flow cytometric analysis of mitochondrial mass, mitochondrial membrane potential and ROS.** Melanomas were generally dissociated enzymatically as described above. Equal numbers of dissociated cells (500,000–2,000,000) from subcutaneous tumours, Ficoll-depleted blood, or metastatic nodules were loaded with dyes to assess mitochondrial mass, mitochondrial membrane potential, and ROS levels. The dyes that were used to assess these parameters were all obtained from Life Technologies. We stained the dissociated cells for 20–45 min at 37 °C with 5 µM Mitotracker Green, Mitotracker DeepRed, CellROX Green, or CellROX DeepRed in HBSS-free (Ca<sup>2+</sup>- and Mg<sup>2+</sup>-free) to assess mitochondrial mass, mitochondrial membrane potential, mitochondrial and cytoplasmic ROS, respectively. For each indicator, staining intensity per cell was assessed by flow cytometry in live human melanoma cells (positive for human HLA and dsRed and negative for DAPI and mouse CD45/CD31/Ter119).

**In vivo treatment of xenografts with drugs.** All animal experiments were performed according to protocols approved by the Institutional Animal Care and Use Committee at the University of Texas Southwestern Medical Center (protocol 2011-0118). Unless otherwise stated, 100 freshly dissociated melanoma cells were injected subcutaneously into the right flanks of NSG mice. When tumours became palpable, in some experiments mice were injected subcutaneously with NAC (Sigma, 200 mg kg<sup>-1</sup> day<sup>-1</sup> in 200 µl PBS, pH 7.4) or PBS as a control. Mice were injected with their last NAC dose 10 min before being euthanized for endpoint analysis. In experiments where mice received NAC via the drinking water, NAC was dissolved in PBS at 1 mg ml<sup>-1</sup> and the water was changed every 2 days. In other experiments methotrexate (Tocris, 1.25 mg kg<sup>-1</sup> day<sup>-1</sup> in 100 µl PBS) was injected intraperitoneally 5 days per week. Mice that received methotrexate were simultaneously administered thymidine (Sigma, 3 mg per mouse per day in 100 µl PBS) and hypoxanthine (Sigma, 750 µg per mouse per day in 100 µl PBS) to prevent suppression of nucleotide biosynthesis. Tumour growth was monitored weekly with a caliper. Experiments were terminated when any tumour in the cohort reached 2.5 cm in size. At the end of experiments, blood was collected by cardiac puncture. Organs were analysed for micrometastases and macrometastases by bioluminescence imaging and visual inspection.

**NADPH/NADP<sup>+</sup> measurement.** Subcutaneous tumours or metastatic nodules were surgically excised as quickly as possible after euthanizing the mice then melanoma cells were mechanically dissociated and NADPH and NADP<sup>+</sup> were measured using NADPH/NADP Glo-Assay (Promega) following the manufacturer instructions. Luminescence was measured using a FLUOstar Omega plate reader (BMG Labtech). Values were normalized to protein concentration, measured using a bicinchoninic acid (BCA) protein assay (Thermo).

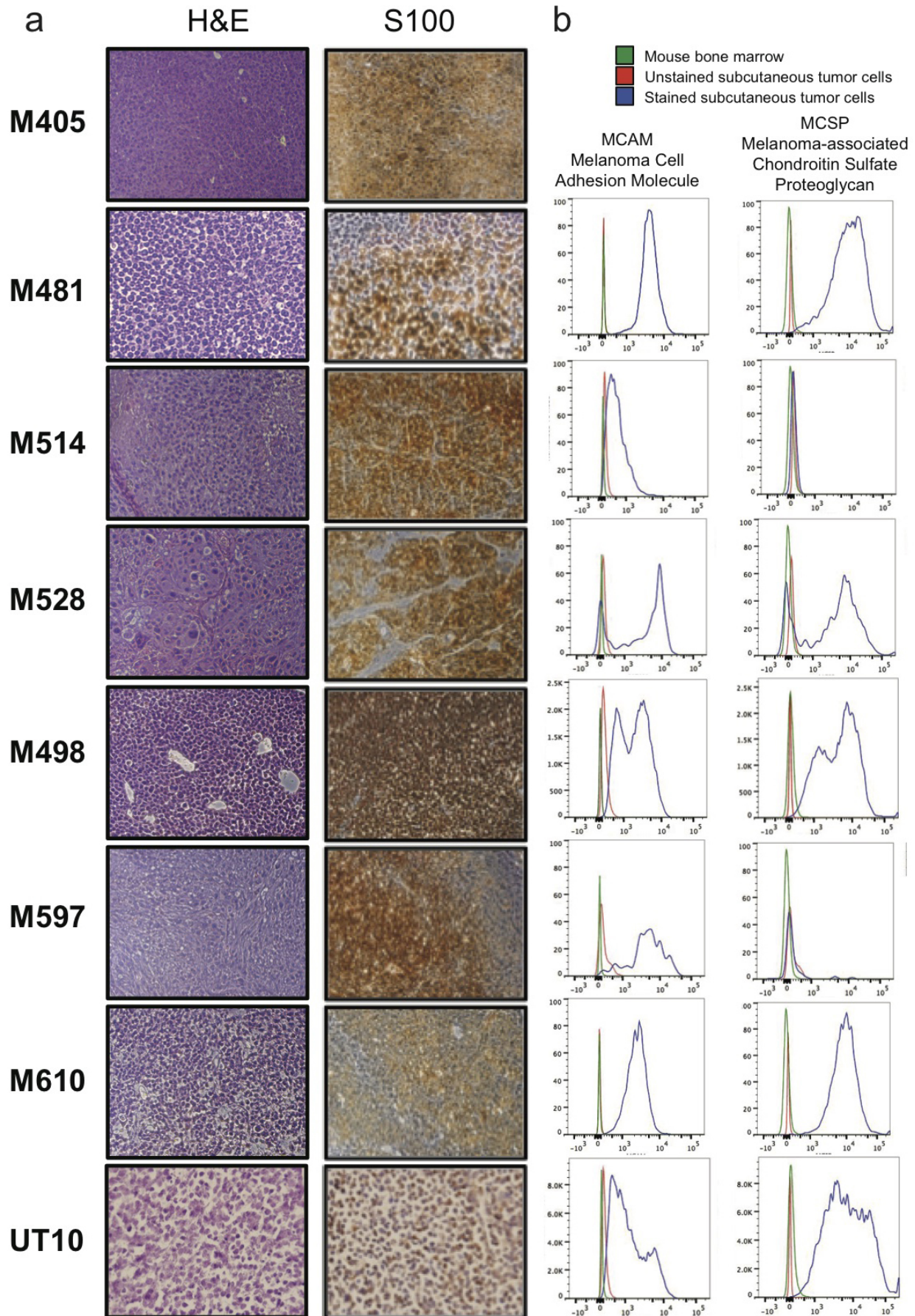
**Western blot analysis.** Tissue lysates were prepared in Kontes tubes with disposable pestles using RIPA Buffer (Cell Signaling Technology) supplemented with phenylmethylsulfonyl fluoride (Sigma), and protease and phosphatase inhibitor cocktails (Roche). The BCA protein assay (Thermo) was used to quantify protein concentrations. Equal amounts of protein (15–30 µg) were separated on 4–20% Tris Glycine SDS gels (BioRad) and transferred to polyvinylidene

difluoride membranes (BioRad). Membranes were blocked for 30 min at room temperature with 5% milk in TBS supplemented with 0.1% Tween20 (TBST) then incubated with primary antibodies overnight at 4 °C. After incubating with horseradish peroxidase conjugated secondary antibodies (Cell Signaling Technology), membranes were developed using SuperSignal West Pico or Femto chemiluminescence reagents (Thermo). Blots were stripped with 1% SDS, 25 mM glycine, pH 2, before re-probing. The following primary antibodies were used for western blot analyses: ALDH1L2 (LifeSpan Bio; LS-C178510), DHFR (LifeSpan Bio; LS-C138829), MTHFR (LifeSpan Bio; LS-C157974), SHMT1 (Cell Signaling; 12612S), SHMT2 (Cell Signaling; 12762S), MTHFD1 (ProteinTech; 10794-1-AP), MTHFD2 (ProteinTech; 12270-1-AP) and aActin (Abcam, ab8227).

**Immunofluorescence staining of frozen tissue sections.** Tissues were fixed in 4% paraformaldehyde for 12 h at 4 °C, and then transferred to 30% sucrose for 24 h for cryoprotection. Tissues were then frozen in OCT. Sections (10 µm) were permeabilized in PBS with 0.2% Triton (PBT), three times for 5 min each, and blocked in 5% goat serum in PBT for 30 min at room temperature. Sections were then stained with primary antibodies overnight: ALDH1L2 (LS-C178510, LifeSpan Bio; 1:50) and S100 (Z0311, Dako, 1:500). The next day, sections were washed in PBS with 0.2% Triton and stained with secondary goat anti-rabbit antibody (Invitrogen) at 1:500 for 30 min in the dark at room temperature. Sections were washed with PBT with DAPI (1:1,000) and mounted for imaging. **Statistical methods.** No statistical methods were used to predetermine sample size. The data in most figure panels reflect several independent experiments performed on different days using melanomas derived from several patients. Variation is always indicated using standard deviation. For analysis of statistical significance, we first tested whether there was homogeneity of variation across treatments (as required for ANOVA) using Levene's test, or when only two conditions were compared, using the F-test. In cases where the variation significantly differed among treatments, the data were log-transformed. If the data contained zero values, 1/2 of the smallest non-zero value was added to all measurements before log transformation. If the data contained negative values, all measurements were log-modulus transformed ( $L(x) = \text{sign}(x) \times \log(|x| + 1)$ ). In the rare cases when the transformed data continued to exhibit variation that significantly differed among treatments, we used a non-parametric Kruskal–Wallis test or a non-parametric Mann–Whitney test to assess the significance of differences among populations and treatments. Usually, variation did not significantly differ among treatments. Under those circumstances, two-tailed Student's *t*-tests were used to test the significance of differences between two treatments. When more than two treatments were compared, a one-way ANOVA followed by Dunnett's multiple comparisons tests were performed. A two-way ANOVA followed by Dunnett's multiple comparisons tests were used in cases where more than two groups were compared with repeated measures. Hierarchical clustering was performed using Euclidean distance in Metaboanalyst<sup>49</sup>.

Mouse cages were randomized between treatments in all *in vivo* experiments (mice within the same cage had to be part of the same treatment). No blinding was used in any experiment. In all xenograft assays we injected 4–8-week-old NSG mice, 5 mice per treatment. Both male and female mice were used. For long-term assays, we injected 10 mice per treatment to account for non-melanoma related deaths (NSG mice are susceptible to death from opportunistic infections). When mice died before the end of experiments due to opportunistic infections the data from those mice were excluded. There were only two experiments in which this occurred. In Fig. 1c, d, 0–4 mice per melanoma line were found dead owing to an opportunistic bacterial infection before termination of the experiment and were excluded from the reported results. In Fig. 2b, 0–3 mice per melanoma line were found dead owing to opportunistic infections, before the first imaging time point after transplantation. These mice were excluded from the reported results.

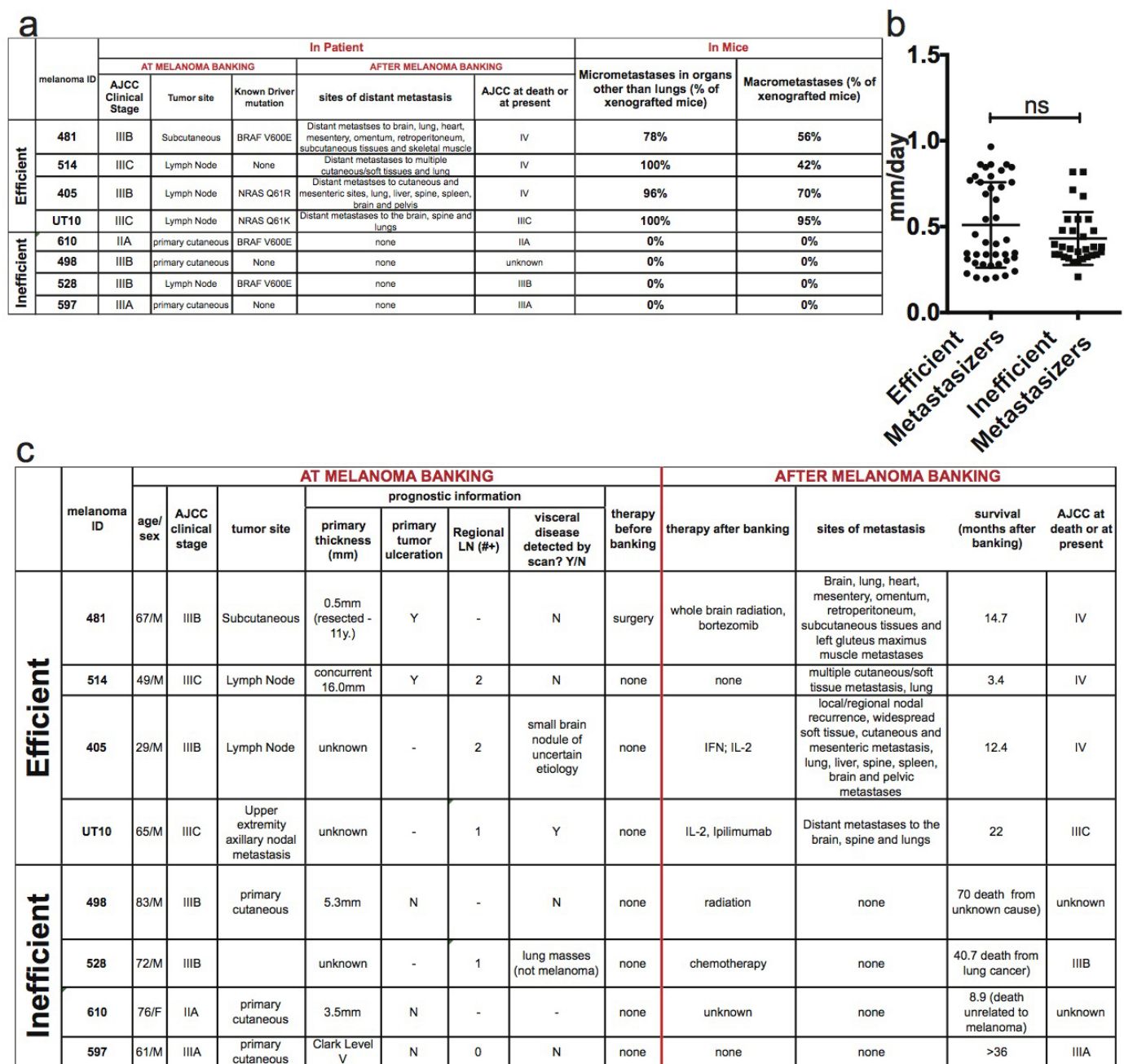
47. Tu, B. P. *et al.* Cyclic changes in metabolic state during the life of a yeast cell. *Proc. Natl Acad. Sci. USA* **104**, 16886–16891 (2007).
48. Mullen, A. R. *et al.* Oxidation of  $\alpha$ -ketoglutarate is required for reductive carboxylation in cancer cells with mitochondrial defects. *Cell Rep.* **7**, 1679–1690 (2014).
49. Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D. & Wishart, D. S. MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic Acids Res.* **40**, W127–W133 (2012).



**Extended Data Figure 1 | Expression of melanoma markers by xenografted melanomas.** **a**, M405, M481, M514, M528, M498, M597, M610 and UT10 tumours were consistently positive for S100, a marker used clinically to diagnose melanoma. **b**, Flow cytometric analysis of xenografted tumour cells that were HLA-ABC<sup>+</sup> and negative for mouse CD31/CD45/Ter119 showed

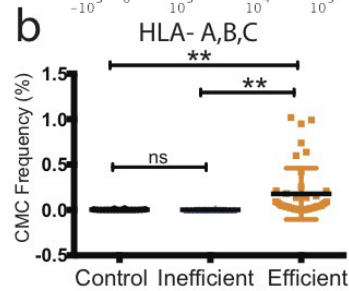
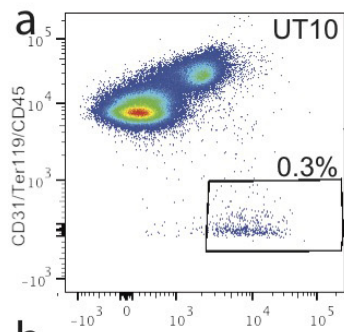
that these cells were usually positive for melanoma cell adhesion molecule (MCAM) and melanoma-associated chondroitin sulphate proteoglycan (MCSP). Both of the tumours that lacked MCSP staining (M514 and M597) were heavily pigmented and expressed other melanoma markers (such as S100 and MCAM).





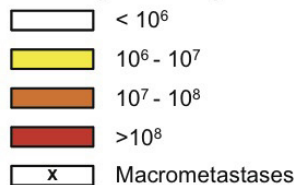
**Extended Data Figure 2 | Clinical data on the melanomas used in this study and summary of their metastatic behaviour in NSG mice.** **a**, Summary of the clinical characteristics of the melanomas used in this study at the time of banking, as well as patient outcome after banking, and metastasis patterns upon transplantation of banked tumours into NSG mice. Melanomas were stratified into efficiently and inefficiently metastasizing melanomas. Efficiently metastasizing melanomas formed distant metastases in patients and in NSG mice, whereas inefficiently metastasizing melanomas did not. The latter group did form micrometastases in the lung, but not outside of the lung in the period of time it took for subcutaneous tumours to grow to 2 cm in diameter (when the mice had to be euthanized in these experiments<sup>27</sup>). Nonetheless,

most of the inefficiently metastasizing melanomas have the ability to form macrometastases if given enough time (data not shown). **b**, Growth rates of subcutaneous tumours in NSG mice after subcutaneous transplantation of 100 cells. Statistical significance was assessed using two-tailed Student's *t*-test. **c**, Clinical characteristics of the patients from whom melanomas were obtained at the time of banking and upon subsequent clinical follow up. The tumours were independently confirmed to be melanomas after xenografting in mice by histological and flow cytometric analysis of melanoma markers (Extended Data Fig. 1) as well as examination by a clinical dermatopathologist.



### Legend for figure 1 (c-f)

Total flux (photon/sec)



### C Efficient Metastasizers

Melanoma	Cell Dose	Lung	Liver	Kidney	Pancreas/Spleen	Gut
M481	10000	X	X			
	10000					
	10000	X	X			
	10000					
	1000					
	1000					
	1000	X	X			
	1000					
	1000	X	X			
	100	X	X			
	100	X	X			
	100	X	X			
M405	1000	X	X	X	X	X
	1000	X	X	X	X	X
	100					
	100					
	100	X	X	X		
	100	X	X	X		
	100	X	X	X		
	100	X	X	X		
	100	X	X	X		
	100	X	X	X		
	100	X	X	X		
	100	X	X	X		
M514	1000		X	X	X	X
	1000		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
	100		X	X	X	X
UT10	1000	X	X	X	X	X
	1000	X	X	X	X	X
	1000	X	X	X	X	X
	1000	X	X	X	X	X
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					

### d Inefficient Metastasizers

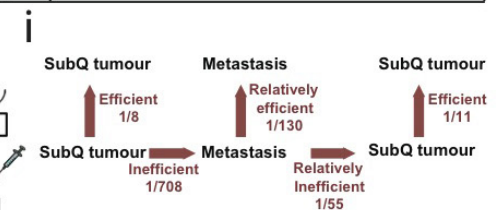
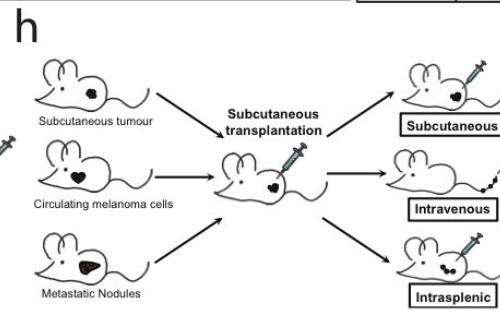
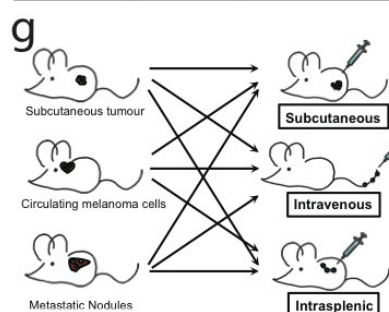
Melanoma	Cell Dose	Lung	Liver	Kidney	Pancreas/Spleen	Gut
M528	10000					
	10000					
	1000					
	1000					
	1000					
	100					
	100					
	100					
	100					
	100	X				
	10					
	10					
M610	1000					
	1000					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
M498/M499	1000					
	1000					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
M597	1000					
	1000					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					

### e Efficient Metastasizers

Melanoma	Cell Dose	Lung	Liver	Kidney	Pancreas/Spleen	Gut
M481	1000		X	X	X	X
	1000		X	X	X	X
	100					
	100		X	X	X	X
	10					
	10	X	X	X	X	X
M405	1000		X			
	100					
	100					
	100					
	100					
	10					
M514	1000		X	X	X	X
	100		X		X	X
	100		X		X	X
	100		X		X	X
	10					
	10					

### f Inefficient Metastasizers

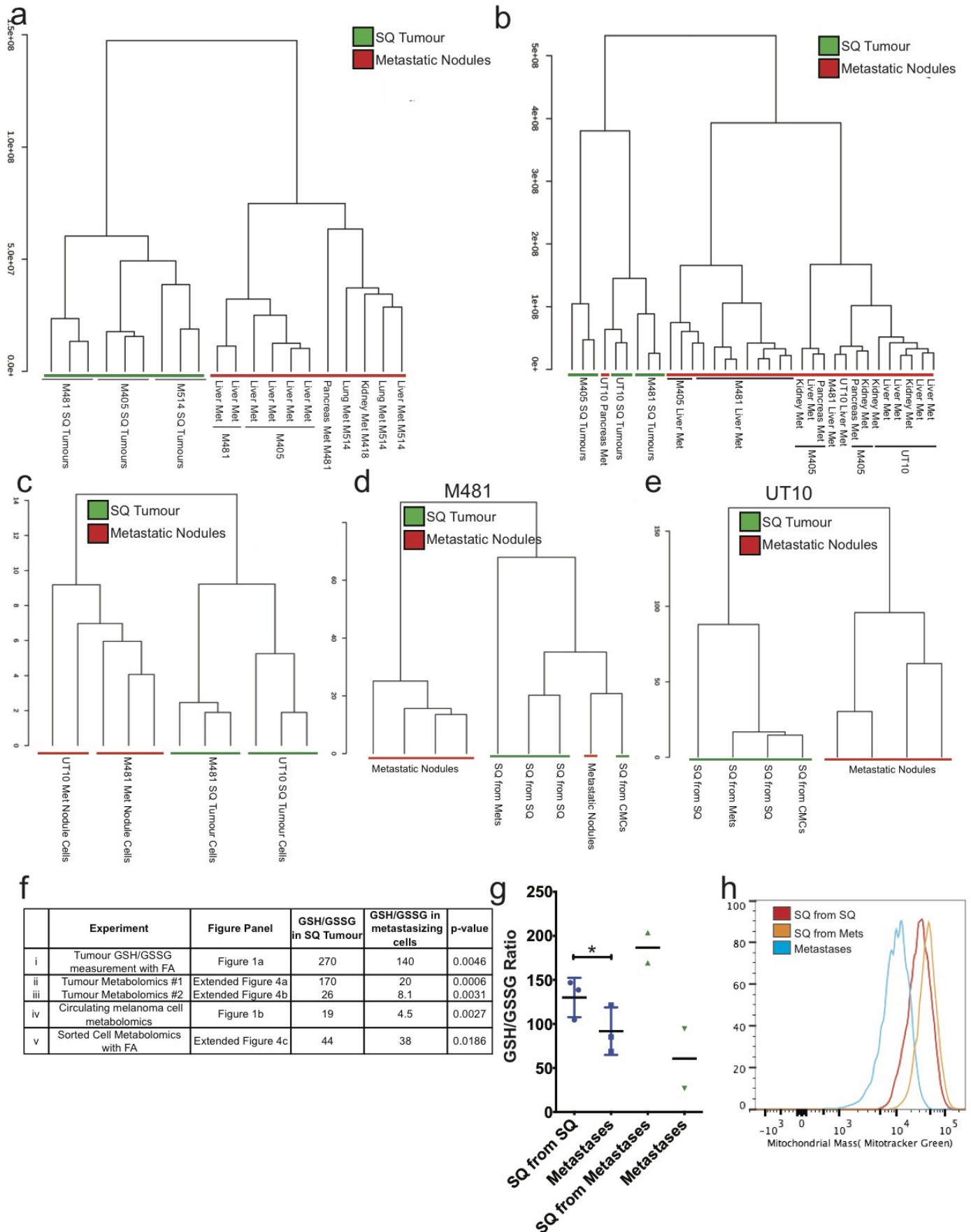
Melanoma	Cell Dose	Lung	Liver	Kidney	Pancreas/Spleen	Gut
M610	1000					
	1000		X			
	1000					
	1000					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
M528	1000					
	1000					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
M498	1000					
	1000					
	1000					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					
	100					



**Extended Data Figure 3 | Barriers to distant metastasis *in vivo*.** **a**, Live human melanoma cells were identified by flow cytometry based on the expression of DsRed (all melanomas in this study stably expressed DsRed) and human HLA and the lack of expression of mouse CD45, CD31 and Ter119 (to exclude mouse haematopoietic and endothelial cells). Human melanoma cells were observed in the blood of NSG mice bearing efficiently metastasizing melanomas. **b**, Mice xenografted with efficiently metastasizing melanomas ( $n = 43$  mice with tumours derived from four patients) had significantly higher frequencies of CMCs in their blood than mice xenografted with inefficiently metastasizing melanomas ( $n = 13$  mice with tumours derived from four patients) or control mice that had not been xenografted ( $n = 18$  mice). Blood

was collected by cardiac puncture. Statistical significance was assessed using ANOVA followed by Tukey's test for multiple comparisons.  $**P < 0.005$ . **c–f**, Bioluminescence analysis of total photon flux (photons  $s^{-1}$ ) from mouse organs after intravenous injection (**c**, **d**) or intrasplenic injection (**e**, **f**) of luciferase-tagged melanoma cells derived from efficiently metastasizing (**c**, **e**) or inefficiently metastasizing (**d**, **f**) melanomas. Each melanoma was derived from a different patient and was studied in an independent experiment. **g**, Schematic of the experiment shown in Table 2a. **h**, Schematic of the experiment shown in Table 2b. **i**, Summary of mean limiting dilution frequencies of tumour-forming cells after subcutaneous, intravenous, or intrasplenic transplantation into NSG mice.

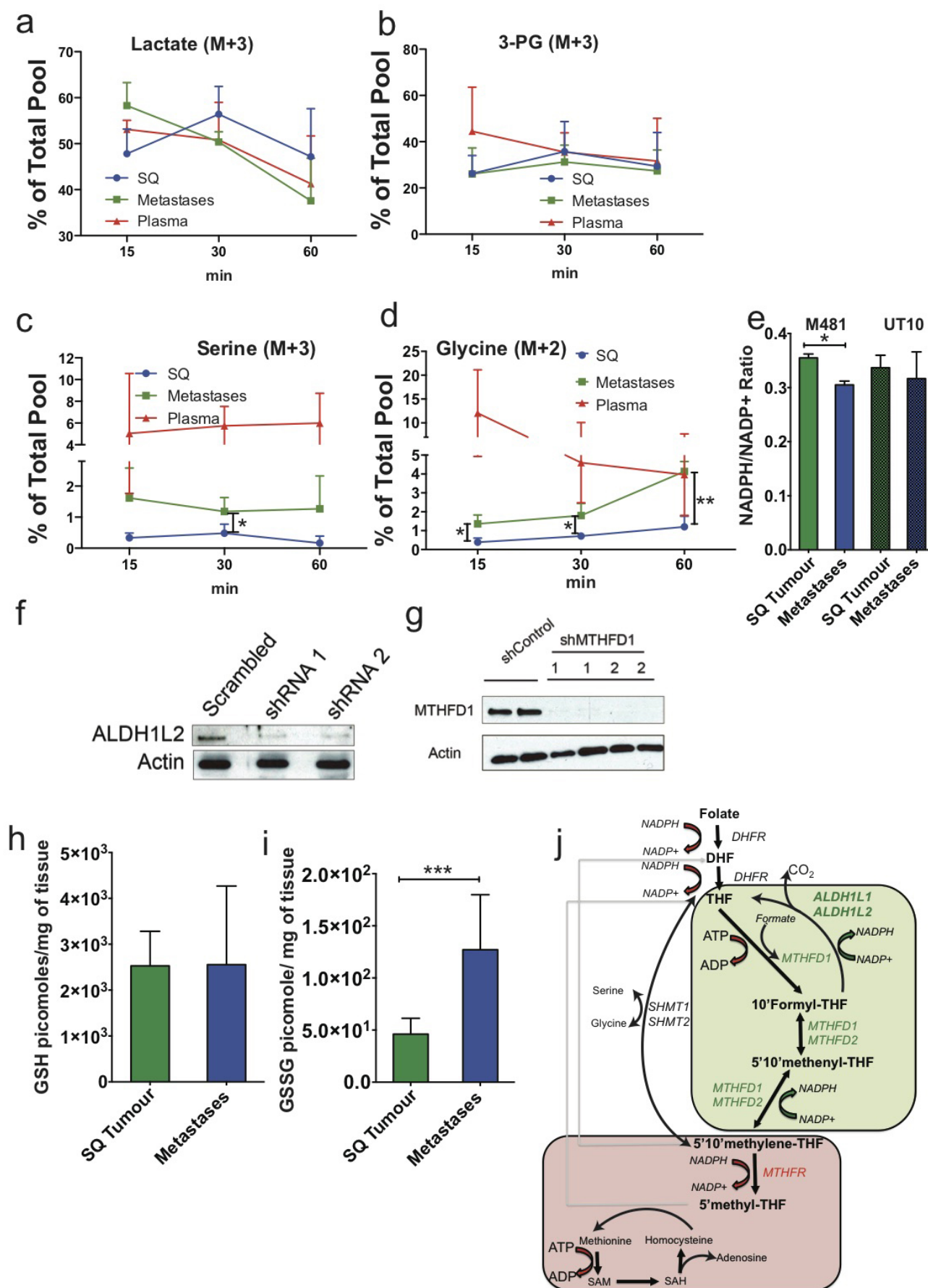




# Extended Data Figure 4 | Unsupervised clustering suggests that melanoma cells undergo reversible metabolic changes during metastasis.

**a, b**, Hierarchical clustering from two independent experiments reflecting subcutaneous tumours and metastatic nodules from the liver, pancreas, lung and kidneys of mice transplanted with melanomas M405, M481 and M514 (**a**) (see Extended Data Table 1 for data on individual metabolites) and subcutaneous tumours and metastatic nodules from the liver, pancreas and kidneys of mice transplanted with melanomas M405, M481 and UT10 ( $n = 2-3$  mice per melanoma in each experiment; see Extended Data Table 2 for data on individual metabolites) (**b**). **c**, Hierarchical clustering of metabolites extracted from flow cytometrically sorted human melanoma cells isolated from subcutaneous tumours or metastatic nodules (UT10, M481,  $n = 3$  mice per melanoma in two independent experiments). **d, e**, Hierarchical clustering of metabolites extracted from subcutaneous tumours and metastatic nodules from mice transplanted subcutaneously with either subcutaneous, circulating or metastatic melanoma cells ( $n = 4$  mice for each melanoma in two independent experiments). **f**, GSH/GSSG ratios from each of the experiments that compared subcutaneous tumours and metastasizing cells. (i) Metabolites were extracted in the presence of 0.1% formic acid to inhibit spontaneous oxidation<sup>42</sup> in two independent experiments comparing subcutaneous and metastatic tumours from mice with three different melanomas in each experiment (M405, M481 and UT10). (ii) and (iii) GSH/GSSG ratios from the experiments

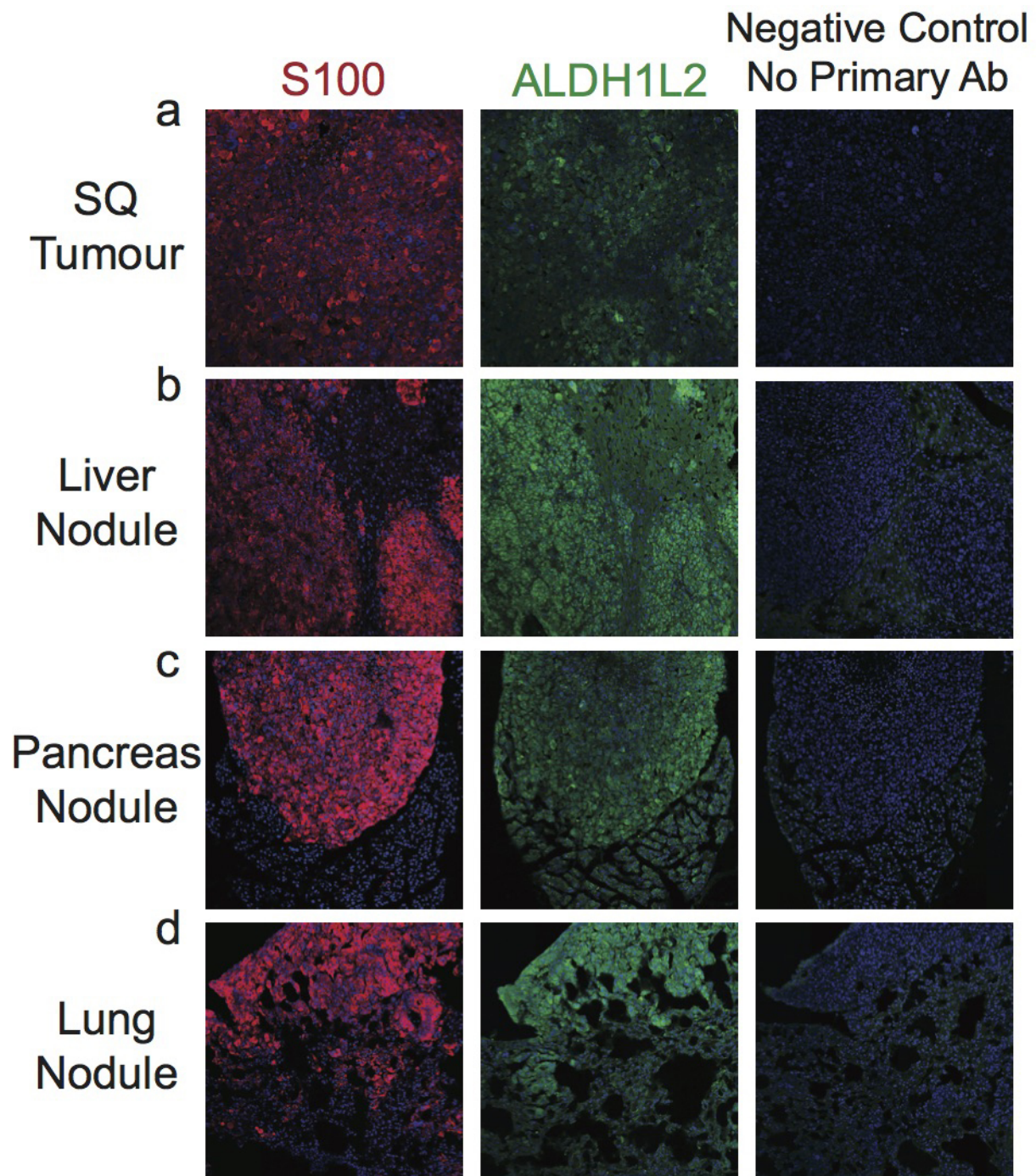
shown in **a** and **b**, respectively. (iv) GSH/GSSG ratios in melanoma cells isolated by flow cytometry from subcutaneous tumours and the blood (circulating melanoma cells) of mice bearing M405 and M481. (v) Metabolites were extracted in the presence of 0.1% formic acid in two independent experiments in which melanoma cells were isolated by flow cytometry from subcutaneous and metastatic tumours (M405 and M481). While the GSH/GSSG ratio was always significantly higher in melanoma cells from subcutaneous tumours as compared to circulating cells or metastatic nodules the ratio varied among experiments as a result of technical differences in cell isolation and metabolite extraction as well as differences in mass spectrometry sensitivity to GSH and GSSG. **g**, GSH/GSSG ratios in subcutaneous tumours that arose from the transplantation of melanoma cells obtained from subcutaneous tumours or metastatic nodules, as well as the metastatic nodules from the same mice (M405;  $n = 2$  to 3 mice per treatment in one experiment). These data suggest that the decline in GSH/GSSG ratio in metastasizing melanoma cells is reversible upon subcutaneous transplantation. **h**, Histogram showing mitochondrial mass in subcutaneous tumour cells that arose from the transplantation of subcutaneous cells (SQ from SQ), subcutaneous tumour cells that arose from the transplantation of metastatic cells (SQ from Mets), and metastatic cells (metastases). These histograms reflect the data shown in Fig. 1g. All error bars represent s.d. Statistical significance was assessed using two-tailed Student's *t*-tests (**f** and **g**). \* $P < 0.05$ .





**Extended Data Figure 5 | Metastatic nodules exhibited increased enrichment of labelled serine and glycine as compared to subcutaneous tumours.** *In vivo* isotope tracing of uniformly  $^{13}\text{C}$ -labelled lactate (M + 3) (a), 3-phosphoglycerate (M + 3) (b), serine (M + 3) (c), and glycine (M + 2) (d) in subcutaneous tumours versus metastatic nodules from the same mice (UT10,  $n = 3$ –4 mice per time point in two independent experiments). The fractional enrichment of labelled lactate, and 3-PG did not significantly differ among plasma, subcutaneous tumours or metastatic tumours at any time point. By contrast, the fractional enrichment of labelled serine and glycine were significantly higher in metastatic as compared to subcutaneous tumours. This is consistent with increased *de novo* serine synthesis in metastatic tumours but could also reflect altered serine/glycine exchange with circulating

serine/glycine pools in metastatic as compared to subcutaneous tumours. e, NADPH/NADP ratios in subcutaneous tumours and metastatic nodules from the same mice shown in Fig. 2d, e. f, g, Western blot of ALDH1L2 (f) and MTHFD1 (g) protein after shRNA knockdown in melanoma cells. Uncropped western blots are shown in Supplementary Fig. 1. h, i, Amount of GSH (h) and GSSG (i) per mg of subcutaneous or metastatic tumour as measured by LC–MS (M405, M481 and UT10,  $n = 2$ –3 mice per melanoma in two independent experiments). All data represent mean  $\pm$  s.d. Statistical significance was assessed using two-tailed Student's *t*-tests (e, h and i) and one-way ANOVAs followed by Dunnett's tests for multiple comparisons (a–d). \* $P < 0.05$ ; \*\*\* $P < 0.0005$ . j, Schematic of the folate pathway including NADPH generating (green box) and NADPH consuming (red box) reactions.



**Extended Data Figure 6 | Immunofluorescence analysis of ALDH1L2 in subcutaneous melanoma as well as metastatic nodules in the liver, pancreas and lung. a–d,** Melanoma cells can be distinguished from host stromal cells based on staining for the melanoma marker, S100.



Extended Data Table 1 | Fold change and statistical significance corrected for multiple comparisons of metabolites detected by LC-MS/MS in the experiment in Extended Data Fig. 2a.

Metabolite	Fold Change of Metastasis/SQ Tumor	FDR	Metabolite	Fold Change of Metastasis/SQ Tumor	FDR
citraconic acid/ketohexanoic acid	12.311	0.054768	tyrosine	0.82787	0.47274
glucarate	11.472	0.13347	leucine	0.82491	0.29642
sedoheptulose 7-phosphate	5.6121	0.084567	citulline	0.82483	0.44062
GABA	4.8296	0.44062	S-lactoylglutathione	0.81622	0.78839
betaine	4.8132	6.12E-05	guanosine	0.81517	0.55117
AICAR	4.5004	0.15975	pyroglutamic acid	0.81294	0.61138
carnitine-C18	4.4942	0.13347	GMP	0.81159	0.64286
spermidine	4.4416	0.016469	3-hydroxybutyrate	0.80538	0.72186
adenosine 5-diphosphoribose	4.1336	0.041494	hydroxyphenylpyruvate	0.798	0.26547
xanthosine	3.9645	0.11311	glycerate	0.77694	0.66141
glutamine	3.9371	0.000419	phenyllactic acid	0.77196	0.02971
hexose-phosphate	3.0656	0.086761	homoserine/threonine	0.7596	0.44062
GSSG	2.8724	0.081173	methionine	0.74125	0.28847
S-methyl-5-thioadenosine	2.8234	0.21534	carnitine	0.73503	0.38386
gluconate	2.7963	0.052588	cADP-ribose	0.73299	0.12096
xanthine	2.4873	0.084567	O-acetylserine	0.72801	0.19553
pyridoxamine	2.4531	0.24821	homocysteine	0.72727	0.16015
spermine	2.3796	0.19553	methylglyoxal	0.72483	0.08875
aminoadipic acid	2.3513	0.44062	creatine	0.72291	0.24821
carnitine-C16	2.3105	0.27824	fumarate	0.7092	0.17141
uridine	2.2733	0.13025	lactate	0.70716	0.13819
citrate	2.1797	0.081173	proline	0.69532	0.27038
methionine sulfoxide	2.0674	0.22847	alpha-ketoisovaleric acid	0.6943	0.14034
ribose 5-phosphate	2.0193	0.40173	cystathionine	0.6937	0.56124
glycerol 3-phosphate	2.0132	0.40173	UMP	0.67863	0.13819
IMP	1.987	0.018053	carbamoylphosphate	0.67623	0.33231
S-adenosylmethionine	1.8423	0.17083	N-acetylglucosamine phosphate	0.65886	0.05477
inosine	1.8253	0.24736	carnitine-C4	0.63354	0.52908
riboflavin	1.7888	0.19553	CMP	0.60248	0.11112
hypoxanthine	1.6282	0.23352	orotate	0.60141	0.15634
acetyllysine	1.5539	0.5435	pipecolic acid	0.6	0.05259
NAD	1.5512	0.47123	fructose	0.59758	0.09166
taurine	1.5405	0.21332	adenine	0.59213	0.03863
thiamine	1.4869	0.47274	hydroxyproline	0.59052	0.00051
aspartate	1.3202	0.68641	GSH	0.58564	0.12364
putrescine	1.3107	0.64286	pantothenic acid	0.57614	0.11592
kynurenine	1.2857	0.56968	cytidine	0.54462	0.05231
dAMP	1.2535	0.61138	phospho-hydroxybenzoate	0.53605	0.31067
thymidine	1.2419	0.6972	phosphocholine	0.53561	0.06755
ornithine	1.2218	0.70841	propionylcholine	0.53061	0.13303
phenylalanine	1.2154	0.49828	1-methyladenosine	0.52411	0.00529
AMP	1.2057	0.69525	glutamate	0.48917	0.01647
S-adenosylhomocysteine	1.1992	0.77036	N-acetylaspargate	0.48911	0.15634
acetylcarnitine	1.1403	0.68641	hydroxyglutarate	0.48629	0.01468
trimethyllysine	1.1395	0.83403	acetylcholine	0.48281	0.03065
carnitine-C8	1.1074	0.78839	shikimic acid	0.46778	0.05604
carnitine-C3	1.0818	0.83403	phosphocreatine	0.46674	0.13819
asparagine	1.0678	0.72161	carnitine-C5	0.45886	0.19875
choline	1.0661	0.83403	alpha-ketoglutarate	0.4554	0.08972
malate	1.0497	0.78839	arginine	0.44217	0.02915
glycerophosphorylcholine	1.0439	0.9381	2-deoxycytidine	0.43995	0.23352
1-methylhistidine	1.0248	0.9381	uracil	0.43712	0.01468
nicotinamide	1.0018	0.99511	5-aminolevulinic acid	0.436	0.00012
glyoxylate	0.98478	0.9381	nicotinate	0.4316	0.19553
dimethylglycine	0.97783	0.91723	lysine	0.40191	6.12E-05
glycine	0.94707	0.91453	1-methylnicotinamide	0.39599	0.33556
succinate	0.94473	0.83403	N,N-dimethylarginine	0.38777	0.00093
alanine	0.92949	0.78839	carnitine-C14	0.3696	0.24821
carnitine-C6	0.92856	0.83403	myo-inositol	0.34547	0.13347
indole	0.92843	0.83403	creatinine	0.29845	0.00047
tryptophan	0.92275	0.80122	ascorbic acid	0.25524	1.57E-06
urea	0.90704	0.66141	XMP	0.15598	0.06755
Guanidoacetic acid	0.88659	0.70463	carnitine-C12	0.15514	0.09166
histidine	0.86073	0.67732	ADP	0.071143	2.54E-08
valine	0.8468	0.66141			



Extended Data Table 2 | Fold change and statistical significance corrected for multiple comparisons of metabolites detected by LC-MS/MS in the experiment shown in Extended Data Fig. 2b.

Metabolite	Fold Change Metastasis/SQ Tumor	FDR	Metabolite	Fold Change Metastasis/SQ Tumor	FDR
citraconic acid/ketohexanoic acid	34.483	0.45284	methionine sulfoxide	0.90754	0.71378
carnitine-C18	10.963	0.02495	citrulline	0.90545	0.69494
NADP	10.741	0.02463	leucine	0.90323	0.39527
adenosine 5-diphosphoribose	6.9242	0.01199	GSH	0.88817	0.69787
carnitine-C16	4.962	0.03569	S-methyl-5-thioadenosine	0.87643	0.60788
hypoxanthine	4.8303	0.01049	putrescine	0.87293	0.3561
pyridoxamine	4.7958	0.02923	cysteine	0.86437	0.7895
betaine	4.7812	0.03627	3-hydroxybutyrate	0.86329	0.5511
S-adenosylhomocysteine	4.3483	0.01917	N-acetylaspargate	0.78325	0.48791
gluconate	4.2564	0.03675	serine	0.78075	0.79908
xanthine	4.1876	0.00192	alanine	0.76685	0.11668
GSSG	4.0565	0.00339	carnitine-C14	0.76026	0.70454
AICAR	3.8575	0.02826	GMP	0.74002	0.18657
xanthosine	3.7994	0.07897	histidine	0.72959	0.0396
adenosine	3.555	0.00983	O-acetylserine	0.72727	0.01269
uridine	3.497	9.33E-05	glutamate	0.72558	0.14273
glutamine	3.318	0.01391	acetylcarnitine	0.68306	0.00192
aconitate	3.2415	0.04627	2-deoxycytidine	0.67868	0.29585
sedoheptulose 7-phosphate	2.9576	0.19981	ascorbic acid	0.67473	0.28456
NAD	2.9139	0.00151	carnitine-C12	0.67254	0.5056
riboflavin	2.8081	0.0222	lactate	0.65954	0.00047
adenosine	2.7184	0.00181	choline	0.65842	0.00389
guanosine	2.4328	0.01321	acetylcholine	0.6573	0.13515
ZMP	2.4059	0.12097	creatinine	0.6543	0.16525
inosine	2.3785	0.14043	γ-aminobutyric acid	0.64139	0.2191
IMP	2.3314	0.01049	glucose/fructose	0.63849	0.00925
deoxyinosine	2.2956	0.20059	homoserine/threonine	0.62876	0.02462
uridine	2.2759	0.01549	dimethylglycine	0.62167	0.00068
SAICA	2.0448	0.13492	carnitine-C4	0.59432	0.16312
malate	1.9365	1.81E-06	N-acetylglucosamine phosphate	0.59273	0.00095
aminoadipic acid	1.9042	0.01321	glucuronate	0.57813	0.00338
carnitine-C8	1.8744	0.0222	UMP	0.57702	0.0163
kynurenine	1.869	0.00672	uridine 5-diphosphoglucuronic acid	0.57518	0.02632
dAMP	1.802	0.00099	phenyllactic acid	0.57402	0.00024
fructose 6-phosphate	1.7965	0.26537	N,N-dimethylarginine	0.5534	0.00047
thymidine	1.6716	0.27532	propionylcholine	0.55055	0.00853
cADP-ribose	1.5868	0.0663	allantoin	0.54186	2.44E-05
cytidine	1.5195	0.21432	urea	0.52852	7.29E-07
carnitine-C6	1.512	0.08539	1-methylnicotinamide	0.51941	0.02008
S-lactoylglutathione	1.46	0.6708	5-aminolevulinic acid	0.51688	8.12E-05
glycerol 3-phosphate	1.4129	0.42166	deoxyribose phosphate	0.50607	0.01321
7-methylguanosine	1.4076	0.55385	uracil	0.50113	0.06104
glyoxylate	1.3983	0.00203	carnitine	0.49567	1.66E-07
S-adenosylmethionine	1.3577	0.15136	taurine	0.46303	2.17E-05
glycerophosphorylcholine	1.335	0.16611	cytosine	0.45951	6.21E-07
tryptophan	1.318	0.19633	proline	0.4591	4.23E-06
1-methyladenosine	1.2813	0.03412	carnitine-C5	0.44249	0.00483
AMP	1.2804	0.43896	ornithine	0.43836	0.00505
citrate	1.2214	0.67146	1-methylhistidine	0.4023	2.31E-06
nicotinamide	1.2063	0.62079	uracil	0.3904	0.01116
succinate	1.196	0.50829	creatine	0.37961	1.26E-07
hydroxyphenylpyruvate	1.1937	0.26901	trimethyllysine	0.37893	7.58E-05
phenylalanine	1.1838	0.1462	arginine	0.36929	0.00047
carnitine-C3	1.1524	0.54363	CMP	0.36814	3.24E-09
tyrosine	1.1166	0.50877	phosphocholine	0.36461	7.43E-09
methionine	1.0869	0.70454	α-ketoglutarate	0.35719	0.02197
2-aminooctanoic acid	1.0803	0.71185	spermidine	0.35496	0.03329
adenine	1.0676	0.74371	acetyllysine	0.35375	3.01E-05
asparagine	1.0509	0.69494	phosphocreatine	0.34497	0.00048
pipecolic acid	1.0216	0.96295	agmatine	0.32422	8.26E-06
thiamine	1.0179	0.97565	hydroxyglutarate	0.30783	3.24E-09
ribose 5-phosphate	1.0132	0.97843	lysine	0.27436	3.91E-10
homocysteic acid	1.0076	0.97843	phosphoenolpyruvic acid	0.27364	1.99E-08
fumarate	1.0022	0.9897	XMP	0.2693	4.33E-05
indole	1.0017	0.9897	cystathionine	0.26823	0.02462
pantothenic acid	0.97947	0.9674	glycine	0.24469	3.32E-09
homocysteine	0.97252	0.96295	carbamoylphosphate	0.17283	8.74E-09
Guanidoacetic acid	0.97007	0.79908	myo-inositol	0.14292	5.08E-07
valine	0.94431	0.69728	cadaverine	0.06894	0.09355
orotate	0.93458	0.90099	nicotinate	0.066878	0.12347

Extended Data Table 3 | Treatment with the antioxidant NAC increased tumour formation by intravenously transplanted melanoma cells.

**a**

Efficient Metastasizers					
		Cell Dose	Mice Injected	Mice that formed tumours	Frequency of tumour-forming cells
<b>Total</b>	With NAC	1000	15	13	1/443
		100	15	4	
	Without NAC	1000	15	3	1/4983
		100	15	0	

\*  
\*  
\***b**

Inefficient Metastasizers					
		Cell Dose	Mice Injected	Mice that formed tumours	Frequency of tumour-forming cells
<b>Total</b>	With NAC	10000	5	2	1/16157
		1000	10	1	
		100	10	0	
	Without NAC	10000	5	0	-
		1000	10	0	
		100	10	0	

Limiting dilution analysis of tumour formation by intravenously transplanted efficiently metastasizing (**a**, M405, M481 and UT10) or inefficiently metastasizing (**b**, M528 and M597) melanoma cells. In half the mice, the cells were pre-treated with NAC before transplantation and the mice were maintained on water supplemented with NAC after transplantation ( $n = 5$  mice/treatment/melanoma for a total of 110 mice in five independent experiments). Statistical significance was assessed by a Chi-square test using ELDA software<sup>46</sup>. \*\*\*\* $P < 0.00005$ .

**Extended Data Table 4 |** Fold increase in metastatic nodules as compared to subcutaneous tumours of transcripts involved in antioxidant responses and NADPH generation.

Gene Name	Gene	Metastatic/Subcutaneous (Fold Change)	p-value*
GCLC	Glutamate-cysteine ligase (catalytic subunit)	1.20	0.560
GCLM	Glutamate-cysteine ligase (modifier subunit)	0.86	0.300
GSS	Glutathione synthase	0.90	0.240
SOD1	Superoxide dismutase 1	1.09	0.355
SOD2	Superoxide dismutase 2	0.86	0.454
SOD3	Superoxide dismutase 3	0.64	0.189
CAT	Catalase	1.09	0.548
GPX1	Glutathione Peroxidase 1	1.30	0.117
GPX2	Glutathione Peroxidase 2	1.00	0.998
GPX3	Glutathione Peroxidase 3	0.40	0.048
GPX4	Glutathione Peroxidase 4	1.07	0.459
GPX7	Glutathione Peroxidase 7	0.83	0.183
GPX8	Glutathione Peroxidase 8	0.87	0.631
GSR	Glutathione Reductase	1.12	0.372
HMOX1	Heme Oxygenase 1	0.86	0.186
HMOX2	Heme Oxygenase 2	1.41	0.008
TXN	Thioredoxin 1	1.16	0.297
TXN2	Thioredoxin 2	1.15	0.480
TXNRD1	Thioredoxin reductase 1	0.97	0.859
TXNRD2	Thioredoxin reductase 2	0.91	0.701
TXNRD3	Thioredoxin reductase 3	1.15	0.176
PRDX1	Peroxiredoxin 1	1.67	0.334
PRDX2	Peroxiredoxin 2	0.94	0.289
PRDX3	Peroxiredoxin 3	1.31	0.013
PRDX4	Peroxiredoxin 4	1.13	0.075
PRDX5	Peroxiredoxin 5	1.28	0.032
PRDX6	Peroxiredoxin 6	1.05	0.172
NFE2L2	Nuclear factor, erythroid 2-like 2	0.90	0.106
G6PD	Glucose-6-phosphate Dehydrogenase	2.20	0.156
PGD	Phosphogluconate dehydrogenase	0.98	0.758
ME2	Malic Enzyme 2	1.05	0.865
CBS	Cystathione beta Synthase	1.20	0.469

Data are based on RNA sequencing (RNA-seq) analysis of two subcutaneous tumours and two metastatic nodules for each of M405 and M481.



# *BCL11A* enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis

Matthew C. Canver<sup>1\*</sup>, Elenoe C. Smith<sup>1\*</sup>, Falak Sher<sup>1\*</sup>, Luca Pinello<sup>2\*</sup>, Neville E. Sanjana<sup>3\*</sup>, Ophir Shalem<sup>3</sup>, Diane D. Chen<sup>1</sup>, Patrick G. Schupp<sup>1</sup>, Divya S. Vinjamur<sup>1</sup>, Sara P. Garcia<sup>2</sup>, Sidinh Luc<sup>1</sup>, Ryo Kurita<sup>4</sup>, Yukio Nakamura<sup>4,5</sup>, Yuko Fujiwara<sup>1,6</sup>, Takahiro Maeda<sup>7</sup>, Guo-Cheng Yuan<sup>2</sup>, Feng Zhang<sup>3§</sup>, Stuart H. Orkin<sup>1,6§</sup> & Daniel E. Bauer<sup>1§</sup>

**Enhancers, critical determinants of cellular identity, are commonly recognized by correlative chromatin marks and gain-of-function potential, although only loss-of-function studies can demonstrate their requirement in the native genomic context. Previously, we identified an erythroid enhancer of human *BCL11A*, subject to common genetic variation associated with the fetal haemoglobin level, the mouse orthologue of which is necessary for erythroid *BCL11A* expression. Here we develop pooled clustered regularly interspaced palindromic repeat (CRISPR)-Cas9 guide RNA libraries to perform *in situ* saturating mutagenesis of the human and mouse enhancers. This approach reveals critical minimal features and discrete vulnerabilities of these enhancers. Despite conserved function of the composite enhancers, their architecture diverges. The crucial human sequences appear to be primate-specific. Through editing of primary human progenitors and mouse transgenesis, we validate the *BCL11A* erythroid enhancer as a target for fetal haemoglobin reinduction. The detailed enhancer map will inform therapeutic genome editing, and the screening approach described here is generally applicable to functional interrogation of non-coding genomic elements.**

Enhancers are classically described as distal genetic elements that positively regulate gene expression in an orientation-independent manner in ectopic heterologous gain-of-function expression experiments<sup>1</sup>. These elements coordinate when, where and how genes are expressed. Enhancer sequences bind transcription factors and are correlated with specific chromatin features including reduced DNA methylation, characteristic histone modifications, heightened chromatin accessibility, long-range promoter interactions, and bidirectional transcription. Recent chromatin mapping has demonstrated the abundance of distal regulatory elements bearing an enhancer signature<sup>2–4</sup>.

The biological importance of enhancers is underscored by gene expression studies showing the predictive power of enhancer profile on lineage-specific programs<sup>5–7</sup>. Highly marked and clustered enhancers (for example, so-called strong, stretch, or super-enhancers) are particularly suggestive of cellular identity and may help to infer lineage-specific regulatory factors<sup>8–10</sup>. Genome-wide association studies reveal enrichment of trait-associated variants in sequences bearing lineage-restricted enhancer signatures<sup>4,8,11,12</sup>. Enhancers display signs of evolutionary constraint as well as heightened turnover with evidence of positive selection<sup>13–16</sup>.

Despite their importance, enhancers are typically defined by criteria unrelated to *in situ* functional requirement. Advances in putative enhancer mapping, as well as large-scale oligonucleotide synthesis, facilitate enhancer reporter assays on a massively parallel scale, allowing a systematic evaluation of the functional significance of enhancer sequences<sup>17,18</sup>. Nonetheless, ectopic heterologous enhancer assays cannot address the necessity of an element in its native chromatin

environment. The growing appreciation of the nonrandom distribution of distal elements both with respect to the linear genome and within the three-dimensional nuclear environment emphasizes the importance of studying enhancers by perturbing their endogenous condition<sup>10,19</sup>.

Insightful observations have been made by mutagenizing enhancers using traditional molecular genetic approaches<sup>20,21</sup>; however, the low throughput of these classical methods constrains their widespread application. Furthermore, the elevated turnover of many enhancer sequences between species may limit the ability to derive conclusions from nonhuman organisms regarding human gene regulation. Advances in genome editing technology make practical the facile modification of the human genome<sup>22,23</sup>. High-throughput Cas9-mediated functional genomics studies have revealed novel genes required for various biological processes<sup>24–27</sup>. Genome editing is likewise suitable for the study of non-coding genetic elements such as enhancers, although these experiments have previously been conducted at low throughput<sup>28–30</sup>.

## Human composite enhancer

Recently, we observed that common genetic variants associated with fetal haemoglobin (HbF;  $\alpha_2\gamma_2$ ) level and  $\beta$ -haemoglobin disorder clinical severity mark an adult-developmental-stage- and erythroid-lineage-specific intronic enhancer of *BCL11A* (ref. 28), a validated repressor of HbF and therapeutic target for  $\beta$ -haemoglobin disorders<sup>28,31–33</sup>. This composite human enhancer is composed of three DNase I hypersensitive sites (DHSSs), termed h+55, h+58 and h+62 on the basis of distance in kilobases (kb) from the transcriptional start site (TSS)<sup>28</sup>. The most highly

<sup>1</sup>Division of Hematology/Oncology, Boston Children's Hospital, Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts 02115, USA. <sup>3</sup>Broad Institute of MIT and Harvard, McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences and Department of Biological Engineering, MIT, Cambridge, Massachusetts 02142, USA. <sup>4</sup>Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki 305-0074, Japan. <sup>5</sup>Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan. <sup>6</sup>Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA. <sup>7</sup>Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.

trait-associated haplotype is defined by two single nucleotide polymorphisms (SNPs): rs1427407 within h+62 and rs7606173 within h+55 (Extended Data Fig. 1a). Previously, we showed that this enhancer possessed ectopic erythroid-restricted, adult-stage-specific enhancer activity<sup>28</sup>. Moreover, the mouse orthologue of the composite enhancer—defined by primary sequence homology, shared erythroid enhancer chromatin signature and syntenic position relative to coding sequences—was shown to be required for *BCL11A* expression and embryonic globin gene repression in a mouse erythroid cell line but dispensable in a mouse B-lymphoid cell line<sup>28</sup>.

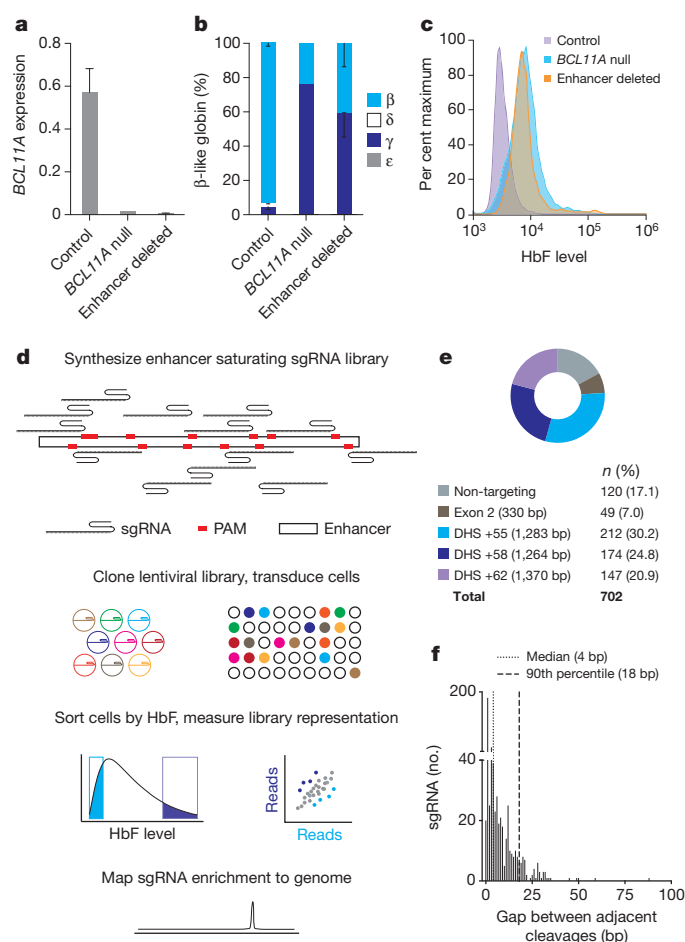
To evaluate the requirement for human *BCL11A* enhancer sequences, we used HUDEP-2 cells, an immortalized human CD34<sup>+</sup> haematopoietic stem and progenitor cell (HSPC)-derived erythroid precursor cell line that expresses *BCL11A* and predominantly  $\beta$ - rather than  $\gamma$ -globin<sup>34</sup>. We used the CRISPR-Cas9 nuclease system to generate clones of HUDEP-2 cells with deletion of the 12-kb *BCL11A* composite enhancer by introduction of a pair of chimaeric single guide RNAs (sgRNAs). Enhancer deletion resulted in near-complete loss of *BCL11A* expression and induction of  $\gamma$ -globin and HbF protein to similar levels as cells with *BCL11A* knockout (Fig. 1a–c), consistent with the possibility that these sequences could serve as targets for therapeutic genome editing for HbF reinduction for the  $\beta$ -haemoglobinopathies<sup>35</sup>. Although targeted deletions by paired double strand breaks (DSBs) may be achieved by genome editing, competing genomic outcomes include local insertion/deletion (indel) production at each cleavage site as well as inversion of the intervening segment<sup>22,23,36–38</sup>.

### Tiled pooled enhancer editing *in situ*

We hypothesized that composite enhancers may be composed of a functional hierarchy with essential and dispensable constituent components. A functional hierarchy might enable enhancer disruption by a single DSB at a critical region followed by non-homologous end joining (NHEJ) repair with indels. In fact, the hypothesis that a prevalent mechanism of trait associations is enhancer variation rests on the premise that single nucleotide changes themselves may substantively modulate enhancer function. Therefore, we reasoned that a tiling set of sgRNAs could uncover critical enhancer regions by disruption of nearly all sequences within an enhancer on the basis of the typical outcome of Cas9 cleavage and NHEJ repair, an indel spectrum with frequent deletions of up to 10 base pairs (bp) from the cleavage position<sup>22,23,36,38,39</sup>.

We designed all possible sgRNAs within the human *BCL11A* composite enhancer DHSs (Fig. 1d, e) as restricted only by the presence of the *Streptococcus pyogenes* Cas9 (SpCas9) NGG protospacer adjacent motif (PAM) sequence, which restricts cleavage at an average 1/8 frequency at each genomic position<sup>22,39</sup>. The NGG-PAM-restricted sgRNAs had a median gap between adjacent genomic cleavages of 4 bp and 90th percentile of 18 bp (Fig. 1f), which suggested that this strategy could approach saturation mutagenesis *in situ*. We included non-targeting sgRNAs as negative controls as well as sgRNAs tiling exon 2 of *BCL11A* as positive controls (Fig. 1e). The library was successfully cloned to a lentiviral vector. The basic experimental schema was to transduce HUDEP-2 cells with the lentiviral library at low multiplicity such that nearly all selected cells contained a single integrant (Fig. 1d). After expansion, differentiation, sorting by HbF level, genomic DNA isolation and deep sequencing of integrated sgRNAs, an HbF enrichment score was calculated for each sgRNA by comparing its representation in HbF-high and HbF-low pools (see Supplementary Information and Extended Data Fig. 2 for additional technical details).

We mapped the HbF enrichment score of each sgRNA to its predicted position of genomic cleavage (Fig. 2a). Most enhancer targeting sgRNAs showed no significant enrichment or depletion from the HbF-high pool. The enriching sgRNAs co-localized to discrete genomic positions. For example, we observed a cluster of sgRNAs at h+62 with modest enrichment, a cluster at h+55 with moderate



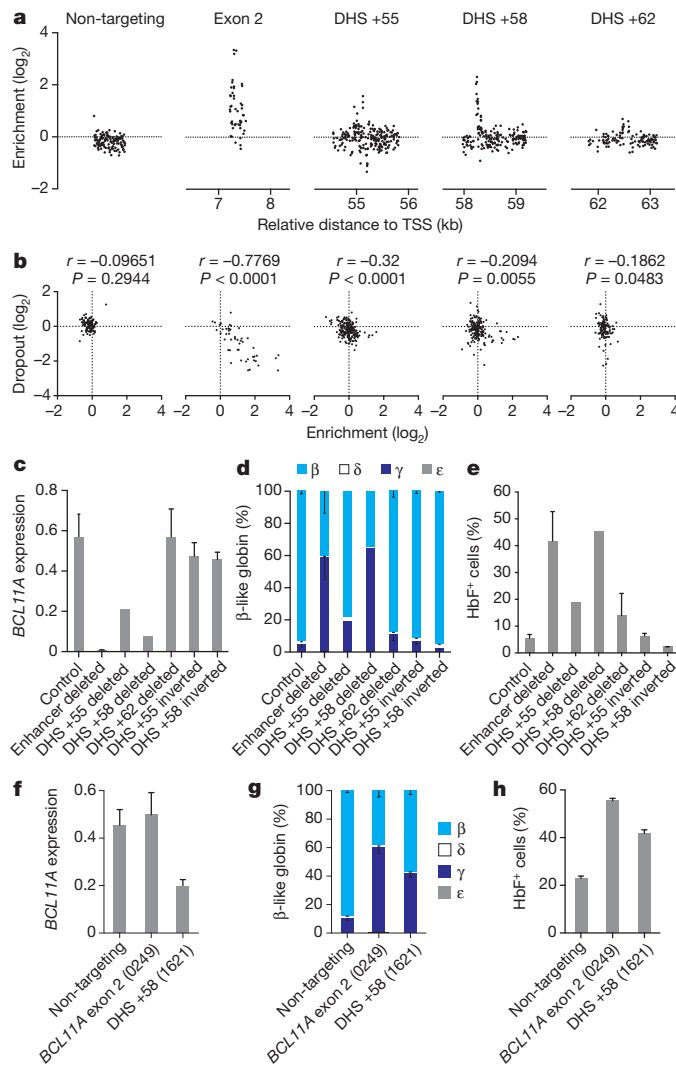
**Figure 1 | Tiled pooled *in situ* CRISPR-Cas9 *BCL11A* enhancer screen.**

**a–c**, Deletion of the human composite *BCL11A* enhancer in HUDEP-2 cells demonstrates its necessity for *BCL11A* expression (normalized to GAPDH), repression of  $\gamma$ -globin mRNA, and repression of HbF; control clones,  $n = 4$ ; *BCL11A* null,  $n = 1$ ; enhancer deleted,  $n = 3$ ; error bars show s.e.m. **d**, Workflow of CRISPR-Cas9 enhancer screen showing library synthesis, delivery and analysis. **e**, Human NGG PAM sgRNA library distribution. **f**, Gaps between adjacent genomic cleavages for NGG PAM sgRNAs targeting *BCL11A* exon 2, h+55, h+58 and h+62.

enrichment (as well as adjacent clusters with depletion), and a cluster at h+58 with marked enrichment. Of note, we observed ten sgRNAs at h+58 with cleavage positions within 42 bp each with HbF enrichment scores exceeding 0.99, the median enrichment score of *BCL11A* exon-2-targeting sgRNAs.

Exon-2-targeted sgRNAs showed a linear correlation between HbF enrichment and cellular dropout, suggesting that sgRNAs that result in complete knockout of *BCL11A* lead to a reduced rate of cell accumulation inseparable from HbF derepression (Fig. 2b). In contrast, the sgRNAs at h+58 associated with marked HbF enrichment showed blunted impact on dropout (Fig. 2b). This finding could be consistent with a low residual level of *BCL11A* adequate to promote cellular accumulation but inadequate to suppress HbF.

To corroborate these findings, we introduced two sgRNAs to the HUDEP-2/Cas9 cells to produce targeted deletion or inversion clones<sup>36</sup>. Deletion of h+58 phenocopied deletion of the composite enhancer and deletion of h+55 had moderate effect (while deletion of h+62 showed a non-significant trend towards a modest effect), consistent with the magnitude of top-scoring and co-localizing sgRNAs from the screen (Fig. 2a, c–e). Inversion of the h+58 or h+55 sites had no significant effect on gene expression, demonstrating that the *BCL11A* enhancer functions in an orientation-independent manner *in situ*, consistent with the classic enhancer definition<sup>1</sup> (Fig. 2c–e).

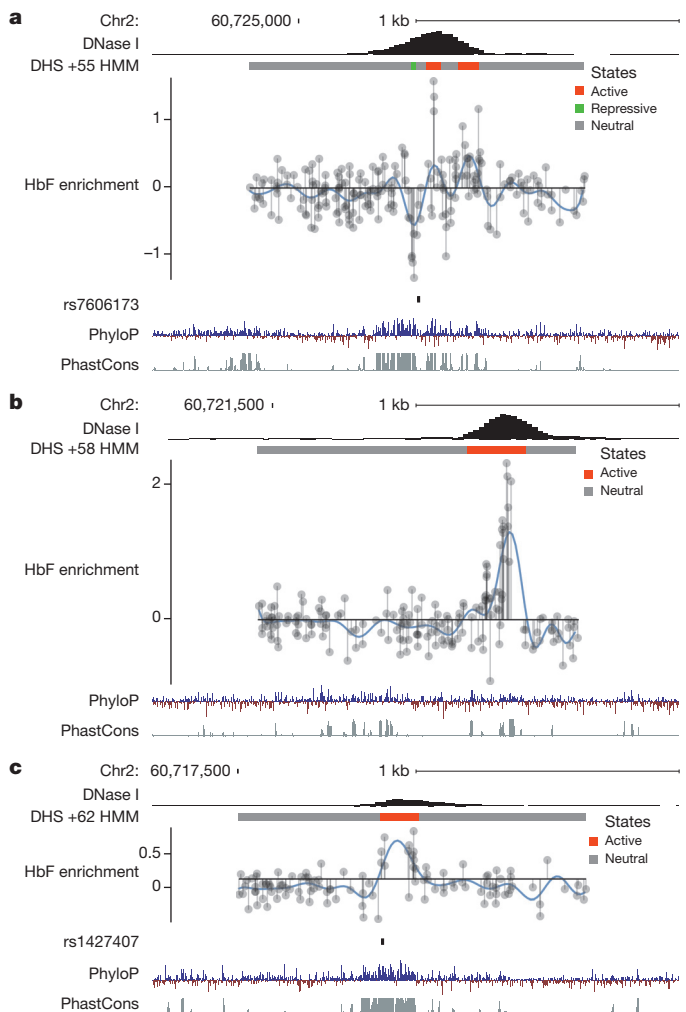


**Figure 2 | Functional mapping of the *BCL11A* enhancer.** **a**, Mapping sgRNA HbF enrichment scores relative to genomic cleavage positions. Non-targeting sgRNAs are pseudo-mapped with 5-bp spacing. **b**, Correlation between cellular dropout and HbF enrichment scores. **c–e**, *BCL11A* expression normalized to GAPDH (**c**),  $\beta$ -like globin expression (**d**), and  $HbF^+$  fraction (**e**) in HUDEP-2 cells with deletion or inversion of individual DHSs; control clones,  $n = 4$ ; enhancer deleted,  $n = 3$ ; +55 deleted,  $n = 1$ ; +58 deleted,  $n = 1$ ; +62 deleted,  $n = 5$ ; +55 inverted,  $n = 3$ ; +58 inverted,  $n = 2$ . **f–h**, *BCL11A* expression normalized to GAPDH (**f**),  $\beta$ -like globin expression (**g**), and  $HbF^+$  fraction (**h**) in primary human erythroid precursors transduced with Cas9 and individual sgRNAs;  $n = 3$ . Error bars represent s.e.m. (**c**, **d**, **f**, **g**) or s.d. (**e**, **h**).

To validate the findings from the HUDEP-2 cells, the top-scoring enhancer-targeting sgRNA from the screen (number 1621 at h+58) was tested in primary human erythroblasts by lentiviral transduction of human CD34<sup>+</sup> HSPCs exposed to *ex vivo* erythroid culture conditions. Consistent with the screen results, sgRNA-1621 resulted in downregulation of *BCL11A* expression and corresponding upregulation of  $\gamma$ -globin expression and increase in  $HbF^+$  cells (Fig. 2f–h). Notably, sgRNA-1621 did not alter surface marker profile, enucleation frequency, or cellular morphology (Extended Data Fig. 3c). Together, these results suggest proof-of-principle of an individual sgRNA targeting a non-coding element for therapeutic genome editing of  $\beta$ -haemoglobin disorders.

### Primate-specific enhancer sequences

We applied a hidden Markov model (HMM) to the sgRNA enrichment score data to infer functionally important sequences within each

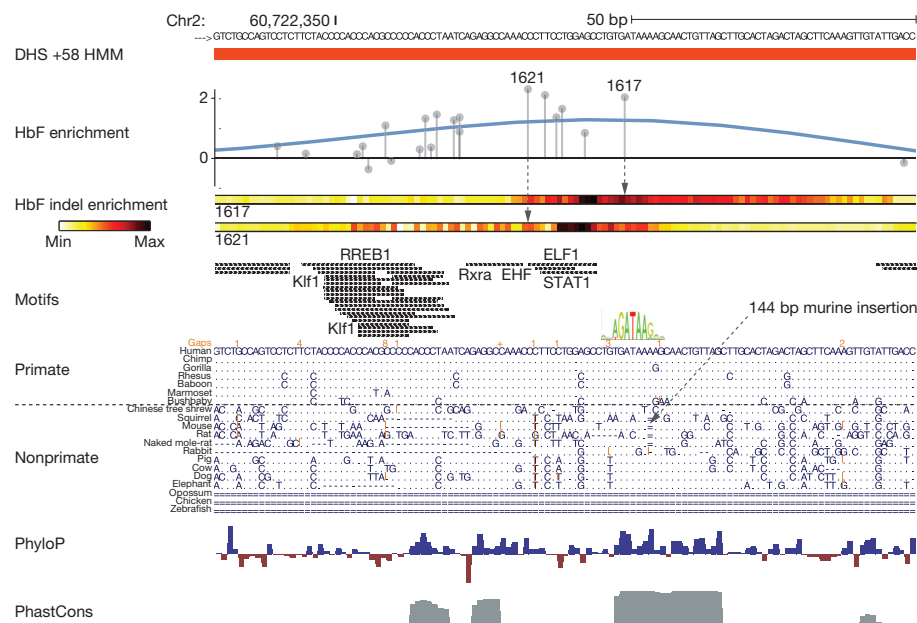


**Figure 3 | Inferred functional enhancer states relative to genomic features.** **a–c**, HMM segmentation of functional enhancer states. HbF enrichment scores are shown throughout DHSs h+55, h+58 and h+62 by grey lines and circles, with the blue line representing smoothed enrichment score. DNase I sequencing from primary human erythroblasts<sup>28</sup>. PhyloP (scale from  $-4.5$  to  $4.88$ ) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 100 vertebrates are shown. Positions of SNPs rs7606173 and rs1427407 are denoted, which together define the haplotype most highly associated with HbF level<sup>28</sup>.

DHS (Extended Data Fig. 4a). This model defined three functional states—active, repressive and neutral—based on likelihood to encompass sequences that positively, negatively and neutrally regulate target gene expression, respectively. The model identified functional states within each DHS (Fig. 3a–c). At each of the three DHSs, the active states were precisely located at regions with the highest degree of DNase I sensitivity.

The overall sequence conservation at the h+58 active region appears both less intense and less distinct from flanking sequences as compared to those of h+62 and h+55 (Fig. 3a–c). The top-scoring sgRNAs in the screen co-localize to 42 bp within h+58 (Fig. 4 and Extended Data Fig. 5b). The third-highest-scoring enhancer-targeted sgRNA (sgRNA-1617) mapped directly onto an apparent GATA1 motif, although below a genome-scale significance threshold ( $P = 3.74 \times 10^{-4}$ ). The mouse orthologous sequence has a GATA1 motif  $P$  value only modestly higher than has the human ( $P = 4.33 \times 10^{-4}$ ). This GATA1 motif appears to have relatively high vertebrate conservation, with exact human sequence identity in rabbits, pigs, dogs and elephants. The top-scoring sgRNA (sgRNA-1621) mapped to a position 15 bp from this GATA1 motif (Fig. 4). An additional four sgRNAs mapping between sgRNA-1621





**Figure 4 | Primate-specific *BCL11A* enhancer functional core.** DHS h+58 functional core defined by maximal HbF enrichment score and active HMM state. HbF enrichment scores are shown by grey lines and circles. HbF indel enrichment per nucleotide is based on amplicon genomic sequencing of sorted cells exposed to either sgRNA-1617 or -1621. No common SNPs (minor allele frequency >1%) are present at this region. JASPAR motifs ( $P < 10^{-4}$ ) are depicted in black with selected motifs annotated by transcription factor based on known erythroid-specific function or genomic position. Gata1 motif LOGO at sgRNA-1617 cleavage position as described in text. Orthologous sequences are listed from representative primates and nonprimates of distributed phylogeny. PhyloP (scale from -4.5 to 4.88) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 100 vertebrates are shown.

and sgRNA-1617 each had substantially elevated HbF enrichment scores. Underlying these sgRNAs were additional predicted motifs (that is, RXRA, EHF, ELF1 and STAT1). Although these sequences showed a high level of conservation among primates, they showed high degeneracy among nonprimate vertebrates (Fig. 4).

We tested the pattern of mutations observed upon treatment of cells with either sgRNA-1621 or sgRNA-1617 by deep sequencing. Each of these sgRNAs is sufficient to substantially induce HbF in human erythroid cells (Fig. 2h and Extended Data Fig. 3a, b). We sorted cells exposed to Cas9 and these sgRNAs into HbF-high and HbF-low pools. We determined the indel spectrum in each population by deep sequencing (Extended Data Fig. 4b). As expected, we observed indels clustering around the predicted cleavage positions. By comparing the per nucleotide indel ratio between cells from the HbF-high and HbF-low pools, we calculated a relative indel enrichment across the sequencing amplicon. Notably, both sgRNAs yielded maximal HbF indel enrichment not precisely at the expected cleavage position but offset at shared intervening sequences (Fig. 4). These sites of maximal HbF mutation enrichment mapped to 7 bp directly overlapping predicted motifs (Fig. 4). Taken together, these data suggest that a conserved GATA1 motif scoring below the prediction threshold adjacent to primate-specific sequences form the core of an enhancer essential for human erythroid *BCL11A* expression and HbF repression.

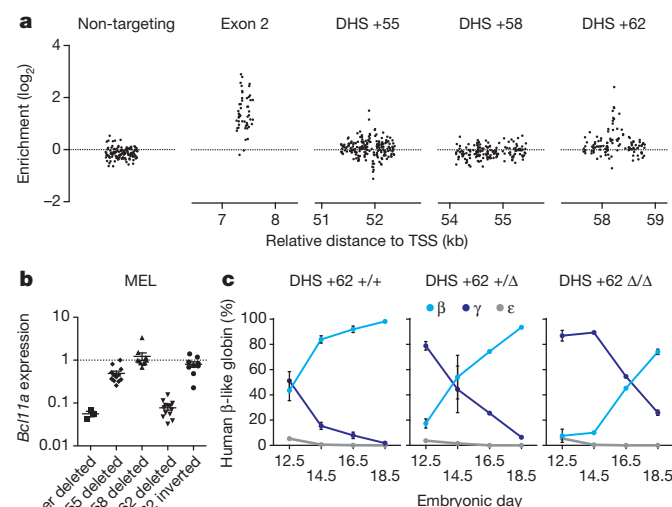
### Mouse enhancer dissection

To test functional conservation of the *BCL11A* enhancer, we examined the orthologous mouse *Bcl11a* enhancer in greater detail. Erythroid DNase I sensitivity is observed at those sequences homologous to h+55 and h+62 but not h+58 (Extended Data Fig. 6a), consistent with the reduced sequence homology within the h+58 active region (Fig. 3a–c). We performed a pooled CRISPR enhancer saturating mutagenesis screen in MEL  $\epsilon\gamma$ :mCherry reporter cells, similar to the human screen described above (Extended Data Figs 6 and 7, and Supplementary Information).

Upon mapping the sgRNA cleavage positions to the genome, we again observed that the majority of enhancer-targeting sgRNAs demonstrated no significant  $\epsilon\gamma$  enrichment or depletion. We observed co-localization of sets of sgRNAs with  $\epsilon\gamma$  enrichment (Fig. 5a). There was a similar complex pattern at the m+55 orthologue as at h+55, with adjacent regions with enriching and depleting sgRNAs from the high- $\epsilon\gamma$ :mCherry pool at the DHS core. At the m+58 orthologue we

did not observe any evidence of  $\epsilon\gamma$  enriching or depleting sgRNAs. At the m+62 orthologue there was a marked peak, with five sgRNAs with  $\epsilon\gamma$  enrichment scores exceeding 1.30, the median enrichment score of *Bcl11a* exon-2-targeting sgRNAs (Fig. 5a). This potent impact of the m+62 orthologue was in contrast to the modest impact of individual sgRNAs or DHS deletion at h+62.

We used pairs of sgRNAs in the presence of Cas9 to produce MEL clones with deletions of various substituent elements at the *Bcl11a* enhancer (Fig. 5b). Deletion of the DNase-insensitive m+58 orthologue had no apparent effect on *Bcl11a* expression, consistent with the pooled screen result. Deletion of the m+55 orthologue led to an approximately twofold reduction in *Bcl11a* expression (mean



**Figure 5 | Functional sequence requirement at the mouse *Bcl11a* erythroid enhancer for *in vivo* haemoglobin switching.** **a**, Mapping sgRNA  $\epsilon\gamma$  enrichment scores to genomic cleavage positions. Non-targeting sgRNAs were pseudo-mapped with 5 bp spacing. **b**, *Bcl11a* expression in mouse erythroid clones with deletion or inversion of individual DHSs relative to non-deleted controls. **c**, Transgenic human  $\beta$ -like globin expression in  $\beta$ -YAC/+62 deletion mice. For +/+ , +/- and -/-: at E12.5,  $n = 5$ , 11 and 3 embryos, respectively; at E14.5,  $n = 2$ , 3 and 4; at E16.5,  $n = 2$ , 4 and 1; at E18.5,  $n = 3$ , 1 and 3. Error bars represent s.e.m.

residual level 49%,  $P < 0.0001$ ), whereas deletion of the m+62 orthologue approached deletion of the entire composite enhancer in terms of reduction in *Bcl11a* expression (mean residual levels of 8% ( $P < 0.0001$ ) and 6% ( $P < 0.0001$ ), respectively, Fig. 5b; also see Supplementary Information and Extended Data Figs 8 and 9). In addition, clones in which the m+62 orthologue was inverted showed no change in *Bcl11a* expression, suggesting that the mouse enhancer, like the human enhancer, functions independently of orientation *in situ* (Figs 2c–e and 5b).

### Erythroid-restricted function *in vivo*

To substantiate the importance of the m+62 orthologue for *Bcl11a* expression and to validate *BCL11A* enhancer disruption as a therapeutic strategy, we generated mice deficient for the *Bcl11a* m+62 orthologue. We used the same Cas9 and paired sgRNA deletion strategy in mouse embryonic stem cells, from which we derived mice with germline transmission of the enhancer deletion. Previous studies have demonstrated essential roles for *Bcl11a* in structural development of the central nervous system and B lymphopoiesis<sup>40–42</sup>. Notably, unlike conventional *Bcl11a* knockouts, which die hours after birth, m+62 orthologue deletion mice were born healthy at expected Mendelian ratios (Extended Data Fig. 10a). The m+62 orthologue deletion mice also showed normal frequencies of B-cell progenitors in the fetal liver and mature B lymphocytes in the adult peripheral blood (Extended Data Fig. 10b, c). Other haematopoietic lineages were also observed at wild-type frequencies (Extended Data Fig. 10c). *Bcl11a* expression was unperturbed in the brain or sorted B cell precursors from embryonic day 16.5 (E16.5) embryos (Extended Data Fig. 10d). In contrast, there was substantial reduction in *Bcl11a* levels in sorted E16.5 erythroid precursors (26% residual,  $P < 0.05$ ; Extended Data Fig. 10d).

The m+62 orthologue deletion mice were bred to mice transgenic for the human  $\beta$ -globin cluster ( $\beta$ -YAC) to model the role of *BCL11A* in haemoglobin switching<sup>43</sup>. Unlike its fetal-stage expression in humans, in the mouse fetal liver transgenic human  $\gamma$ -globin is subject to intense repression (like an embryonic globin). *Bcl11a* is required for this early murine silencing of transgenic  $\gamma$ -globin at E14.5, although even in the absence of *Bcl11a*,  $\gamma$ -globin is ultimately repressed<sup>32,33</sup>. Fetal livers were evaluated between days E12.5 and E18.5 to monitor haemoglobin switching. Repression of human  $\gamma$ -globin and activation of human  $\beta$ -globin was markedly delayed in the m+62 orthologue deleted mice (Fig. 5c). Heterozygous mice showed an intermediate  $\gamma$ -globin derepression phenotype, underscoring the dose-dependent inverse relationship between *BCL11A* and HbF level. These results indicate that targeting the erythroid enhancer of *Bcl11a* *in vivo* results in erythroid-specific disruption of *Bcl11a* expression and relaxed repression of  $\gamma$ -globin, unaccompanied by the obvious neurological and immunological toxicities seen in the *Bcl11a* conventional knockout context.

### Discussion

We employed a novel application of CRISPR-Cas9 genome editing, saturating mutagenesis of non-coding elements *in situ*, to provide an important insight into the organization and function of the *BCL11A* erythroid enhancer. Traditional tests of enhancer function rely on ectopic heterologous reporter assays and/or correlative biochemical features. Genome editing allows facile evaluation of the requirement of enhancer sequences within their endogenous chromatin context for appropriate gene regulation. As shown here, high-resolution, high-throughput pooled tiling sgRNA screening reveals underlying enhancer sequence requirements approaching nucleotide resolution. A limitation to the resolution of this approach is the availability of NGG PAM sequences in a given region. We did not observe efficient editing by SpCas9 with NAG-restricted sgRNAs (Extended Data Figs 2h and 6j). Recent studies have identified Cas9 orthologues and variants restricted by alternative PAM sequences, each capable of efficient genome editing<sup>44–46</sup>. This increased targeting range of Cas9

could allow increased resolution for *in situ* mutagenesis, particularly at sequences with paucity of NGG motifs. Alternatively, approaches reliant on homology-directed repair<sup>47</sup> could offer nucleotide-resolution functional mutagenesis of non-coding sequences, although issues of efficiency, fidelity and quantitative sensitivity would need to be considered. We suggest that our tiled pooled CRISPR screening approach could be readily adapted to the functional interrogation of numerous non-coding genomic elements.

In addition, these data demonstrate that apparent sequence conservation at the *BCL11A* enhancer masks underlying functional divergence. The mouse and human *BCL11A* erythroid composite enhancers share primary sequence homology, an erythroid enhancer chromatin signature, and syntenic intronic position relative to coding sequences. Moreover, both are required for erythroid expression of *BCL11A* and repression of embryonic/fetal globin genes. However, our high-resolution CRISPR mutagenesis analysis reveals divergence in the architecture of these enhancers. Of note, human *BCL11A* enforces the  $\gamma$ - to  $\beta$ -globin developmental switch around the time of birth. The timing and nature of these switches and the globin genes themselves are distinct in primates as compared to nonprimate vertebrates that only exhibit a mid-gestation embryonic to adult switch<sup>48</sup>. Therefore, it would seem plausible that critical regulatory mechanisms at *BCL11A* might differ between species (also see Supplementary Information).

The haemoglobin disorders are among the most common Mendelian inherited human conditions. The level of HbF is a key modifier of clinical severity of these diseases and *BCL11A* is the chief regulator of HbF level<sup>48</sup>. Naturally occurring genetic variation at the *BCL11A* enhancer is well-tolerated and associated with HbF level and  $\beta$ -haemoglobin disorder clinical severity. The work presented here offers a framework for therapeutic genome editing of the *BCL11A* enhancer for  $\beta$ -haemoglobin disorders. Enhancer disruption by individual sgRNAs in primary erythroid precursors results in substantial HbF induction. This approach may mitigate erythroid-specific growth disadvantages of complete *BCL11A* loss (Fig. 2b). Furthermore, erythroid enhancer disruption may spare *BCL11A* expression and function in non-erythroid contexts, such as B lymphopoiesis (Extended Data Fig. 10b–d). A challenge for the field is that it is not yet possible to accurately model HbF repression experimentally. However, individuals haploinsufficient for *BCL11A* due to microdeletions exhibit marked neurological deficits, and elevated HbF beyond that seen in homozygotes for high-HbF common enhancer haplotypes<sup>49,50</sup>. Taken together, these data suggest that perturbation of critical sequences within the *BCL11A* enhancer defined here may result in HbF levels exceeding a clinical threshold required to ameliorate the  $\beta$ -haemoglobin disorders.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 April; accepted 25 August 2015.

Published online 16 September 2015.

1. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a  $\beta$ -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
2. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
3. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **488**, 75–82 (2012).
4. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
5. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
6. Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
7. Xu, J. *et al.* Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev. Cell* **23**, 796–811 (2012).
8. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).

9. Parker, S. C. J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl Acad. Sci. USA* **110**, 17921–17926 (2013).
10. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
11. Paul, D. S. *et al.* Maps of open chromatin guide the functional follow-up of genome-wide association signals: Application to hematological traits. *PLoS Genet.* **7**, e1002139 (2011).
12. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
13. Hardison, R. C. Variable evolutionary signatures at the heart of enhancers. *Nature Genet.* **42**, 734–735 (2010).
14. Vierstra, J. *et al.* Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007–1012 (2014).
15. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
16. Pennacchio, L. A. *et al.* *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
17. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnol.* **30**, 271–277 (2012).
18. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nature Biotechnol.* **30**, 265–270 (2012).
19. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).
20. Bender, M., Bulger, M., Close, J. & Groudine, M.  $\beta$ -globin gene switching and DNase I sensitivity of the endogenous  $\beta$ -globin locus in mice do not require the locus control region. *Mol. Cell* **5**, 387–393 (2000).
21. Johnson, K. D. *et al.* Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J. Clin. Invest.* **122**, 3692–3704 (2012).
22. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
23. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
24. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
25. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
26. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnol.* **32**, 267–273 (2014).
27. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
28. Bauer, D. E. *et al.* An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* **342**, 253–257 (2013).
29. Gröschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. *Cell* **157**, 369–381 (2014).
30. Mansour, M. R. *et al.* An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
31. Sankaran, V. G. *et al.* Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322**, 1839–1842 (2008).
32. Sankaran, V. G. *et al.* Developmental and species-divergent globin switching are driven by BCL11A. *Nature* **460**, 1093–1097 (2009).
33. Xu, J. *et al.* Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science* **334**, 993–996 (2011).
34. Kurita, R. *et al.* Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS ONE* **8**, e59890 (2013).
35. Hardison, R. C. & Blobel, G. A GWAS to therapy by genome edits? *Science* **342**, 206–207 (2013).
36. Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289**, 21312–21324 (2014).
37. Mandal, P. K. *et al.* Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell* **15**, 643–652 (2014).
38. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
39. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31**, 827–832 (2013).
40. Liu, P. *et al.* Bcl11a is essential for normal lymphoid development. *Nature Immunol.* **4**, 525–532 (2003).
41. John, A. *et al.* Bcl11a is required for neuronal morphogenesis and sensory circuit formation in dorsal spinal cord development. *Development* **139**, 1831–1841 (2012).
42. Yu, Y. *et al.* Bcl11a is essential for lymphoid development and negatively regulates p53. *J. Exp. Med.* **209**, 2467–2483 (2012).
43. Porcu, B. S. *et al.* The human  $\beta$  globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. *Blood* **90**, 4602–4609 (1997).
44. Ran, F. A. *et al.* *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
45. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered and improved PAM specificities. *Nature* **523**, 481–485 (2015).
46. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10**, 1116–1121 (2013).
47. Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120–123 (2014).
48. Bauer, D. E., Kamran, S. C. & Orkin, S. H. Reawakening fetal hemoglobin: Prospects for new therapies for the beta-globin disorders. *Blood* **120**, 2945–2953 (2012).
49. Basak, A. *et al.* Persistence of fetal hemoglobin and altered neurodevelopment due to BCL11A deletions. *JCI* **125**, 2363–2368 (2015).
50. Funnell, P. W. *et al.* 2p15-p16.1 microdeletions encompassing and proximal to BCL11A are associated with elevated HbF in addition to neurologic impairment. *Blood* **126**, 89–93 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Hughes and D. Higgs for assistance with analysis of ChIP-seq; R. Mathieu and the Boston Children's Hospital Hematology/Oncology-HSCI Flow Cytometry Research Facility for cell sorting; Z. Herbert and F. Abderazzaq at the Dana-Farber Cancer Institute Molecular Biology Core Facility and Center for Cancer Computational Biology, respectively, for sequencing; J. Doench for providing TALENs; C. Peng for advice with MEL reporter cell generation; F. Godinho and M. Nguyen for technical help with ESCs and transgenic mice; A. Dass, C. Lin and S. Kamran for general technical assistance; C. Brendel and D. Williams for input regarding lentiviral transduction of HSPCs; J. Desimini for graphical assistance; and J. Xu and G. Lettre for insightful discussions. M.C.C. is supported by F30DK103359-01A1. E.C.S. is supported by a Jane Coffin Childs Memorial Fund for Medical Research Fellowship. L.P. is supported by NHGRI Career Development Award K99HG008399. N.E.S. is supported by a Simons Center for the Social Brain Postdoctoral Fellowship and NIH NHGRI award K99-HG008171. O.S. is supported by a fellowship from the Klarman Family Foundation. S.L. is supported by a Leukemia & Lymphoma Society Fellow Award. T.M. is supported by NIH R01 A1084905. G.-C.Y. is supported by NIH R01HL119099 and R01HG005085. F.Z. is supported by the NIMH (5DP1-MH100706) and NIDDK (5R01-DK097768), a Waterman award from the National Science Foundation, the Keck, McKnight, Damon Runyon, Searle Scholars, Merkin, Vallee, and Simons Foundations, and Bob Metcalfe. S.H.O. is supported by P01HL032262 and P30DK049216 (Center of Excellence in Molecular Hematology). D.E.B. is supported by an NIDDK Career Development Award K08DK093705, Doris Duke Charitable Foundation Innovations in Clinical Research Award (2013137), and Charles H. Hood Foundation Child Health Research Award. Computational tools and instructions for designing CRISPR-Cas9 sgRNA libraries for conducting non-coding screening can be found at the Zhang laboratory website <http://www.genome-engineering.org>.

**Author Contributions** D.E.B. conceived this study. N.E.S., O.S. and F.Z. conceived the pooled non-coding screening strategy using CRISPR-Cas9. M.C.C., N.E.S., O.S., F.Z., S.H.O. and D.E.B. designed and executed the pooled CRISPR screening strategy. E.C.S., F.S., Y.F., S.L., S.H.O. and D.E.B. designed, produced and analysed the transgenic mice. R.K. and Y.N. provided the HUDEP-2 cell line. M.C.C., F.S., T.M., S.H.O. and D.E.B. adapted the HUDEP-2 cell line as a model of globin gene regulation. M.C.C., F.S., D.D.C., P.G.S., D.S.V. and D.E.B. performed all experiments in cell lines. M.C.C., L.P., N.E.S., S.P.G., G.-C.Y., F.Z., S.H.O. and D.E.B. analysed the data. L.P., S.P.G. and G.-C.Y. developed the HMM. M.C.C., S.H.O., and D.E.B. wrote the manuscript with input from all authors.

**Author Information** All reagents described in this manuscript have been deposited with Addgene (<http://www.addgene.org>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.B. ([daniel.bauer@childrens.harvard.edu](mailto:daniel.bauer@childrens.harvard.edu)), S.H.O. ([stuart\\_orkin@dfci.harvard.edu](mailto:stuart_orkin@dfci.harvard.edu)), or F.Z. ([zhang@broadinstitute.org](mailto:zhang@broadinstitute.org)).



## METHODS

No statistical methods were used to predetermine sample size.

**Generation of genomic deletions in HUDEP-2 cells.** HUDEP clone 2 (HUDEP-2) was used as previously described<sup>34</sup>. HUDEP-2 cells were expanded in StemSpan SFEM (Stem Cell Technologies) supplemented with  $10^{-6}$  M dexamethasone (Sigma),  $100 \text{ ng ml}^{-1}$  human stem cell factor (SCF) (R&D), 3 international units (IU)  $\text{ml}^{-1}$  erythropoietin (Amgen), 1% L-glutamine (Life Technologies), and 2% penicillin/streptomycin.  $1 \mu\text{g ml}^{-1}$  doxycycline (Sigma) was included in the culture to induce expression of the human papilloma virus type 16 E6/E7 genes<sup>34</sup>. HUDEP-2 cells were differentiated in Iscove's Modified Dulbecco's Medium (IMDM) (Life Technologies) supplemented with  $330 \mu\text{g ml}^{-1}$  holo-transferrin (Sigma),  $10 \mu\text{g ml}^{-1}$  recombinant human insulin (Sigma),  $2 \text{ IU ml}^{-1}$  heparin (Sigma), 5% human solvent detergent pooled plasma AB (Rhode Island Blood Center), 3 IU  $\text{ml}^{-1}$  erythropoietin,  $100 \text{ ng ml}^{-1}$  human SCF,  $1 \mu\text{g ml}^{-1}$  doxycycline, 1% L-glutamine, and 2% penicillin/streptomycin.

Tandem sgRNA lentiviruses were transduced into HUDEP-2 with stable Cas9 expression (Supplementary Table 1). Bulk cultures were incubated for 7–10 days with  $10 \mu\text{g ml}^{-1}$  blasticidin and  $1 \mu\text{g ml}^{-1}$  puromycin selection to allow for editing. Then bulk cultures were plated clonally at limiting dilution. 96 well plates with greater than 30 clones per plate were excluded to avoid mixed clones. After approximately 14 days of clonal expansion, genomic DNA was extracted using 50  $\mu\text{l}$  QuickExtract DNA Extraction Solution per well (Epicentre). Clones were screened for deletion by conventional PCR with one PCR reaction internal to segment to be deleted (non-deletion band) and one gap-PCR reaction across the deletion junction (deletion band) that would only amplify in the presence of deletion<sup>36</sup>. Biallelic deletion clones were identified as the absence of the non-deletion PCR band and the presence of the deletion PCR band (Supplementary Table 2). Inversion clones were identified as previously described by PCR<sup>36</sup> (Supplementary Table 3). Briefly, inversion clones had one inverted allele and one deleted allele without the presence of non-deletion alleles. In our experience biallelic inversion clones are very rare events<sup>36</sup>. PCR was performed using the Qiagen HotStarTaq 2 $\times$  master mix and the following cycling conditions:  $95^\circ\text{C}$  for 15 min; 35 cycles of  $95^\circ\text{C}$  for 15 s,  $60^\circ\text{C}$  for 1 min,  $72^\circ\text{C}$  for 1 min;  $72^\circ\text{C}$  for 10 min. Alternatively, PCR was also performed using 2 $\times$  Accuprime Supermix II (Life Technologies) with the following cycling conditions:  $94^\circ\text{C}$  for 2 min; 35 cycles of  $94^\circ\text{C}$  for 20 s,  $60^\circ\text{C}$  for 20 s,  $68^\circ\text{C}$  for 1 min  $\text{kb}^{-1}$  of PCR product;  $68^\circ\text{C}$  for 5 min. RNA was extracted from each positive clone using a kit (Qiagen) and quantitative real-time RT-qPCR was performed using iQ SYBR Green Supermix (Bio-Rad). Primers used are found in Supplementary Table 5. Gene expression was normalized to that of *GAPDH*. We isolated four control, one *BCL11A* null, three composite enhancer deleted, one h+55 deleted, one h+58 deleted, five h+62 deleted, three h+55 inverted, and two h+58 inverted clones. The *BCL11A* null clone had a 216 bp interstitial deletion of exon 2, preventing binding of the RT-qPCR primers. All gene expression data reported from these clones represents the mean of at least three technical replicates.

**Design and synthesis of human and mouse lentiviral sgRNA libraries.** Every 20-mer sequence upstream of an NGG or NAG PAM sequence on the plus or minus strand was identified for both the human and mouse orthologous +55, +58 and +62 DHS as well as *BCL11A/Bcl11a* exon 2 (Fig. 1 and Extended Data Figs 2, 6). Relative to the human hg19 reference genome, a reference was used with the following substitutions to approximate a common low-HbF-associated haplotype: rs1427407-G, rs1896293-T, rs6706648-T, rs6738440-G, rs7606173-C. The mouse orthologous sequences to each of the human DHSs were defined by using the liftOver tool of UCSC Genome Browser as previously described<sup>28</sup>. Each of the sgRNA oligos were synthesized as previously described<sup>25,51,52</sup> and cloned using a Gibson Assembly master mix (New England Biolabs) into lentiGuide-Puro (Addgene plasmid ID 52963) which had been BsmBI digested, gel purified, and dephosphorylated. Gibson Assembly products were transformed to electrocompetent cells (E. coli, Lucigen). Sufficient colonies were isolated to ensure  $\sim 90\times$  library coverage for both human and mouse libraries. Plasmid libraries were deep sequenced to  $533\times$  and  $813\times$  coverage for human and mouse libraries, respectively, to confirm representation.

To produce lentivirus, HEK293T cells were cultured with Dulbecco's Modified Eagle's Medium (DMEM) (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Omega Scientific) and 2% penicillin-streptomycin (Life Technologies) in 15 cm tissue culture treated Petri dishes. HEK293T cells were transfected at 80% confluence in 12 ml of media with  $13.3 \mu\text{g}$  psPAX2,  $6.7 \mu\text{g}$  VSV-G, and  $20 \mu\text{g}$  of the lentiviral construct plasmid of interest using  $180 \mu\text{g}$  of linear polyethylenimine (Polysciences). Medium was changed 16–24 h after transfection. Lentiviral supernatant was collected at 48 and 72 h post-transfection and subsequently concentrated by ultracentrifugation (24,000 rpm for 2 h at  $4^\circ\text{C}$  with Beckman Coulter SW 32 Ti rotor).

**Tiled pooled CRISPR-Cas9 screen for *in situ* functional mapping the human *BCL11A* erythroid enhancer.** HUDEP-2 cells with stable Cas9 expression were transduced at low multiplicity with the human sgRNA library lentivirus pool while in expansion medium. Control transductions were performed to ensure transduction rate did not exceed 50%. Cell numbers were maintained throughout the experiment at levels adequate to exceed  $1,000\times$  representation of the library.  $10 \mu\text{g ml}^{-1}$  blasticidin (Sigma) and  $1 \mu\text{g ml}^{-1}$  puromycin (Sigma) were added 24 h after transduction to select for lentiviral library integrants in cells with Cas9. Cells were cultured in expansion media for one week followed by differentiation media for an additional week.

Intracellular staining was performed by fixing cells with 0.05% glutaraldehyde (grade II) (Sigma) for 10 min at room temperature. Cells were centrifuged for 5 min at 600g and then resuspended in 0.1% Triton X-100 (Life Technologies) for 5 min at room temperature for permeabilization. Triton X-100 was diluted with phosphate buffered saline (PBS) with 0.1% BSA and then centrifuged at 600g for 15 min. Cells were stained with anti-human antibodies for HbF (clone HbF-1 with FITC or APC conjugation; Life Technologies) and  $\beta$ -haemoglobin antibody (clone 37-8 with PerCP-Cy5 or PE conjugation; Santa Cruz) for 20 min in the dark. Cells were washed to remove unbound antibody before FACS analysis.  $0.2 \mu\text{g}$  HbF and  $2 \mu\text{g}$  of adult haemoglobin (HbA) ( $\beta$ -haemoglobin) antibodies were used per 5 million cells. Control cells exposed to a non-targeting sgRNA sample and *BCL11A* exon 2 were used as negative and positive controls, respectively, to establish flow cytometry conditions. Populations of cells with the top and bottom 10% of expression of HbF were sorted by FACS.

After sorting the HbF-high and HbF-low pools, library preparation and deep sequencing was performed as previously described<sup>25</sup>. Briefly, genomic DNA was extracted using the Qiagen Blood and Tissue kit. Herculanase PCR reaction (Agilent) using lentiGuide-Puro specific primers (5'-AATGGACTATCATA TGCTTACCGTAACCTTGAAAGTATTTTCG-3' and 5'-CTTTAGTTTGAT GTCTGTTGCTATTATGTCTACTATTCTTTCCC-3') including a handle sequence was performed as follows: Herculanase II reaction buffer (1 $\times$ ), forward and reverse primers (0.5  $\mu\text{M}$  each), dimethyl sulfoxide (DMSO) (8%), deoxynucleotide triphosphates (dNTPs) (0.25 mM each), Herculanase II Fusion DNA Polymerase (0.5 reactions) using the following cycling conditions:  $95^\circ\text{C}$  for 2 min; 20 cycles of  $95^\circ\text{C}$  for 15 s,  $60^\circ\text{C}$  for 20 s,  $72^\circ\text{C}$  for 30 s;  $72^\circ\text{C}$  for 5 min. Multiple reactions of no more than 200 ng each were used to amplify from  $6.6 \mu\text{g}$  gDNA ( $\sim 1\times 10^6$  cell genomes) per pool. Samples were subjected to a second PCR using handle-specific primers<sup>25</sup> to add adaptors and indexes to each sample using the following conditions: Herculanase II reaction buffer (1 $\times$ ), forward and reverse primers (0.5  $\mu\text{M}$  each), dNTPs (0.25 mM each), Herculanase II Fusion DNA Polymerase (0.5 reactions) with the following cycling conditions:  $95^\circ\text{C}$  for 2 min; 25 cycles of  $95^\circ\text{C}$  for 15 s,  $60^\circ\text{C}$  for 20 s,  $72^\circ\text{C}$  for 30 s;  $72^\circ\text{C}$  for 5 min. PCR products were run on an agarose gel and the band of expected size was gel purified. Illumina MiSeq 150 bp paired end sequencing was performed.

sgRNA sequences present in the plasmid pool as well as in the HbF-high and HbF-low pools were enumerated. Guide sequences were mapped to the guides comprising the sgRNA library without allowing mismatches. Total reads were normalized to library sequencing depth. Cellular dropout score was determined by calculating (1) the ratio of normalized reads in the cells at end of experiment (average of reads in the HbF-high and HbF-low pools) to reads in the plasmid pool; (2)  $\log_2$  transformation; and (3) median of biological replicates. HbF enrichment score was determined by calculating (1) the ratio of normalized reads in the HbF-high compared to reads in the HbF-low pools; (2)  $\log_2$  transformation; and (3) median of biological replicates. After exclusion of sgRNAs with dropout scores  $< 2^{-3}$  and NAG PAM sgRNAs, a quantile-quantile plot was made with a line fitted through the first and third quantiles using R software. HbF enrichment scores and cellular dropout scores were compared by Spearman rank correlation. sgRNA sequences were mapped to the human genome (hg19) with cleavage positions set to between positions 17 and 18 given PAM positions 21–23. For visual comparisons to targeting sgRNAs, non-targeting sgRNAs were pseudomapped each separated by 5 bp.

**Validation in primary human CD34<sup>+</sup> HSPCs.** Primary human CD34<sup>+</sup> HSPCs from G-CSF mobilized healthy adult donors were obtained from the Center of Excellence in Molecular Hematology at the Fred Hutchinson Cancer Research Center, Seattle, Washington. CD34<sup>+</sup> HSPCs were subject to erythroid differentiation liquid culture as previously described<sup>53</sup>. Briefly, HSPCs were thawed on day 0 into erythroid differentiation medium (EDM) consisting of IMDM supplemented with  $330 \mu\text{g ml}^{-1}$  holo-human transferrin,  $10 \mu\text{g ml}^{-1}$  recombinant human insulin,  $2 \text{ IU ml}^{-1}$  heparin, 5% human solvent detergent pooled plasma AB, 3 IU  $\text{ml}^{-1}$  erythropoietin, 1% L-glutamine, and 2% penicillin/streptomycin. During days 0–7 of culture, EDM was further supplemented with  $10^{-6}$  M hydrocortisone (Sigma),  $100 \text{ ng ml}^{-1}$  human SCF, and 5 ng  $\text{ml}^{-1}$  human IL-3 (R&D).

During days 7–11 of culture, EDM was supplemented with 100 ng ml<sup>-1</sup> human SCF only. During days 11–18 of culture, EDM had no additional supplements.

HSPCs were transduced with lentiCas9-Blast (Addgene plasmid ID 52962) 24 h after thawing in the presence of 10  $\mu$ M 16,16-dimethylprostaglandin E2 (PGE2; Cayman Chemical). At 48 h after thawing, medium was changed and cells were transduced with lentiGuide-Puro or lentiGuide-Crimson cloned with relevant sgRNA sequence in the presence of 10  $\mu$ M PGE2. Three independent transductions were performed per sgRNA. At 72 h after thawing, medium was changed and HSPCs were selected with 10  $\mu$ g ml<sup>-1</sup> blasticidin and 1  $\mu$ g ml<sup>-1</sup> puromycin or 10  $\mu$ g ml<sup>-1</sup> blasticidin followed by sorting for lentiGuide-Crimson<sup>+</sup> cells on day 16 of culture. Blasticidin and/or puromycin selection occurred from days 3 to 8 of culture.

Differentiation was assessed on day 18 of culture using anti-human antibodies against the transferrin receptor (CD71) (Clone OKT9 with FITC conjugation; eBioscience) and glycophorin A (CD235a) (Clone HIR2 with PE conjugation; eBioscience). Enucleation was assessed using 2  $\mu$ g ml<sup>-1</sup> of the cell-permeable DNA dye Hoescht 33342 (Life Technologies). CD235a<sup>+</sup>Hoescht 33342<sup>-</sup> cells were determined to be enucleated erythroid cells. Cells were intracellularly stained for HbF and HbA on day 18 of culture as described above. 50,000–100,000 cells were centrifuged onto microscope slides at 350 rpm for 4 min. Slides were stained with Harleco May–Grünwald stain (Millipore) for 2 min, Giemsa stain (Sigma) for 12 min, and two water washes for 30 s each. Slides were air dried and then coverslipped using Fisher Chemical Permount Mounting Medium (Fisher). RNA isolation and RT–qPCR was performed as above. Gene expression was normalized to that of *GAPDH*. All gene expression data represent the mean of at least three technical replicates.

PCR primers were designed to amplify the genomic cleavage site for a given sgRNA. Resulting PCR products were subjected to Sanger sequencing. Sequencing traces were used for editing quantification using a previously described publically available tool<sup>54</sup>.

**Computational analysis.** Human erythroid H3K27ac ChIP-seq was obtained from Xu *et al.*<sup>7</sup> and mouse erythroid H3K27ac ChIP-seq was obtained from Kowalczyk *et al.*<sup>55</sup> and Dogan *et al.*<sup>56</sup>. We uniformly processed all the data sets using the same pipeline with the same criteria to call super-enhancers. Specifically, we started from raw reads and realigned each data set with Bowtie2 with the default parameters. We then removed duplicate reads with the Picard Suite. To call the peaks we used MACS2 in the narrow mode. Finally, to call the super-enhancers we used the ROSE algorithm with the default parameters<sup>10</sup>. Using these settings, peaks closer than 12.5 kb are stitched together and then ranked based on the H3K27ac intensity. To assign super-enhancers to genes we used again ROSE with default settings. In particular, the tool reports three categories of genes for each super-enhancer: (1) overlapping genes (genes for which the gene body region overlaps a super-enhancer); (2) proximal genes (genes close to a super-enhancer considering a window of 50 kb); (3) closest gene (closest gene considering its TSS and the centre of the super-enhancer). To generate a Venn diagram of genes for super-enhancer data sets, we used the union of these three gene categories.

HMM segmentation was performed to automatically segment the enrichment score signals into enhancer regions with active, repressive and neutral effect. We designed a HMM with 3 states using the GHMM package (<http://ghmm.sourceforge.net/>). To learn the HMM parameters we used the Baum–Welch algorithm. To find the best segmentation for each region we used the Viterbi algorithm. The emission probability for each state was modelled as a Gaussian distribution and all the possible transitions between states were allowed as shown in Extended Data Fig. 4a. Since the signal was not obtained with a constant genomic resolution, we interpolated and smoothed the signal using a Gaussian kernel over 12 bp and applied the HMM to the smoothed signal. To set the initial parameters, we used the 1%, 50% and 99% percentile of the smoothed signal for the prior of the means of the repressive, neutral and active states, respectively, while the prior for the standard deviation was set to 0.001 for all the three states.

Motif analysis was performed to evaluate the human and mouse enhancer regions for potential binding sites for known transcription factors. We used the FIMO software<sup>57</sup> with a *P*-value threshold of  $<10^{-4}$ . For each region we extracted sequences using the hg19 and mm9 assemblies respectively for human and mouse. The motif database was the latest version of the JASPAR database<sup>58</sup>.

Deep sequencing paired-end reads of genomic amplicons from genome editing target sites were first filtered for reads with PHRED quality score  $<30$ , merged with the FLASH (Fast Length Adjustment of SHort reads) software, and subsequently aligned to a reference amplicon using the needle aligner from the EMBOSS suite (<http://emboss.sourceforge.net/>) to quantify insertions and deletions. Per nucleotide frequency of deletion of a position, insertion directly adjacent to the position, or no mutation at the position was quantitated using CRISPResso (<https://github.com/lucapinello/CRISPResso>).

**Pooled CRISPR-Cas9 screen for high-resolution functional mapping of mouse *Bcl11a* enhancer.** Murine erythroleukaemia (MEL, MEL-745A cl. DS19) cells were cultured in DMEM supplemented with 10% FBS, 1% L-glutamine, and 2% penicillin/streptomycin as previously described<sup>28,36</sup>. Cell lines tested negative for mycoplasma contamination.  $\epsilon$ y-mCherry reporter MEL cells with stable Cas9 expression were transduced at low multiplicity with the mouse sgRNA library lentivirus pool (see Supplementary Information and Extended Data Fig. 6 for additional technical details). Control transductions were performed to ensure that the transduction rate did not exceed 50%. Cell numbers were maintained throughout the experiment at levels adequate to exceed 1,000 $\times$  representation of the library. 10  $\mu$ g ml<sup>-1</sup> blasticidin and 1  $\mu$ g ml<sup>-1</sup> puromycin were added 24 h after transduction to select for lentiviral library integrants in cells with Cas9. Subsequently cells were cultured for 2 weeks. The top and bottom 5% of  $\epsilon$ y-mCherry-expressing cells exposed to the library were sorted by FACS. A non-targeting sgRNA sample was used as a negative control and *Bcl11a* exon 2 as a positive control to establish flow cytometry conditions. After sorting, library preparation and deep sequencing were performed as described for the human library<sup>25</sup>.

sgRNA sequences present in the Hbb- $\epsilon$ y-mCherry-high and Hbb- $\epsilon$ y-mCherry-low pools were enumerated. Cellular dropout and  $\epsilon$ y enrichment scores were calculated analogously to the human screen. sgRNA sequences were then mapped to the mouse genome (mm9).

**Generation of genomic deletions in MEL cells.** Deletions in MEL cells were generated using two sgRNA as previously described<sup>36</sup>. Briefly, sgRNA sequences were cloned into pX330 (Addgene plasmid ID 42230) using a Golden Gate assembly cloning strategy (Supplementary Table 1). MEL cells were electroporated with 5  $\mu$ g of each pX330-sgRNA plasmid and 0.5  $\mu$ g pmax-GFP (Lonza) in BTX electroporation buffer using a BTX electroporator (Harvard Apparatus). Approximately 48 h post-electroporation, the top 1–3% of GFP<sup>+</sup> cells were sorted and plated clonally at limiting dilution. Clones were allowed to grow for 7–10 days. Clones were screened for deletion by conventional PCR using the same strategy as with the HUDEP-2 cells (Supplementary Tables 2 and 4). Inversion clones were identified by PCR as previously described<sup>36</sup> (Supplementary Table 3).

**Generation of genomic deletions in mouse embryonic stem cells (mESCs).** mESCs were maintained on irradiated mouse embryonic fibroblasts (GlobalStem) and cultured in high glucose DMEM supplemented with 15% FBS, 1% L-glutamine, 2% penicillin/streptomycin (Life Technologies), 100  $\mu$ M non-essential amino acids (Life Technologies), 1% nucleosides (Sigma),  $10^{-4}$  M  $\beta$ -mercaptoethanol (Sigma), and  $10^3$  U ml<sup>-1</sup> leukaemia inhibitory factor (Millipore). Cells were passaged using 0.25% trypsin (Life Technologies).

The *Bcl11a* +62 deletion mice were derived from CRISPR-Cas9 modified C99 ES cells. Using Amara ES Cell transfection reagent (Lonza), two million mESCs were electroporated with 2  $\mu$ g of each pX330 plasmid vector containing individual target sequences flanking the +62 site along with 0.5  $\mu$ g of a GFP plasmid. After 48 h, the top 5% of GFP expressing cells were sorted, plated on irradiated fibroblasts and maintained. Individual ES cell colonies were then picked and screened for biallelic deletion using the same strategy as HUDEP-2 and MEL cells<sup>36</sup>. DNA for screening CRISPR-Cas9 modified clones was obtained from gelatin adapted ES cell clones to avoid genomic contamination from the fibroblasts. Correctly targeted clones with greater than 80% normal karyotype were used to generate mice. Clones were injected into embryonic day 3.5 (E3.5) C57BL/6 blastocysts and implanted into pseudo-pregnant females.

The  $\beta$ -YAC mouse line (A20), previously described as containing a transgene encompassing ~150 kb of the human  $\beta$ -globin locus<sup>43</sup>, was used to analyse human globin expression. The mouse line was maintained in a hemizygous state and bred with *Bcl11a* +62 deletion mice. Sufficient matings were established to ensure adequate homozygotes for analysis.

**Mouse cell and tissue analysis.** The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. For developmental haematopoiesis, fetal liver cells were taken at E12.5, E14.5, E16.5 and E18.5 and mechanically dissociated to form single cell suspensions from which RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) and analysed. At E16.5, fetal livers were also stained with CD19-PerCP-Cy5.5 (Clone 1D3; eBioscience), B220-APC (RA3-6B2; Biolegend), CD71-PE (Clone C2; BD Biosciences), and Ter119-FITC (Clone Ter119; BD Biosciences) to isolate B cells (B220<sup>+</sup>CD19<sup>+</sup>) and erythroid cells (Ter119<sup>+</sup>CD71<sup>+</sup>) by FACS for RNA extraction and *Bcl11a* quantification. Additionally, flow cytometry was used to analyse fetal liver from E18.5 embryos. Single cell suspensions were stained with IgM-FITC (Clone II-41; eBioscience), CD19-PerCP-Cy5.5, (Clone 1D3; eBioscience), CD43-PE (Clone S7; eBioscience), AA4.1-PE-Cy7 (Clone AA4.1; BD Biosciences), B220-APC, (RA3-6B2; Biolegend), and DAPI (Invitrogen). For adult haematopoietic assays, peripheral blood was obtained from the tail vein of 4-week-old male and female mice. Blood was collected in EDTA-coated tubes, red



cells removed by 2% dextran (Sigma), residual red cells lysed with ammonium chloride solution (Stem Cell Technologies) and stained with the following anti-mouse antibodies: CD3e-FITC (Clone 145-2C11; Biolegend), CD19-PerCP-Cy5.5 (Clone 1D3; eBioscience), CD71-PE (Clone C2; BD Biosciences), NK1.1-PE-Cy5 (Clone PK136; Biolegend), Ter119-APC (Clone TER-119; Biolegend), Gr-1-eF450 (Clone RB6-8C5; eBioscience), B220-BV605 (RA3-6B2; Biolegend), Mac-1-BV510 (Clone M1/70; Biolegend), and 7-AAD (BD Biosciences). Fetal brain analysis was conducted on whole brains from E16.5 mouse embryos on ice-cold PBS. Tissue was directly lysed into the RLT plus buffer (Qiagen) and total RNA extracted according to manufacturer's instructions provided in the RNeasy Plus Mini Kit. RT-qPCR was performed as above, with gene expression normalized to *Gapdh*. All gene expression data represent the mean of at least three technical replicates. All animal experiments were conducted under the approval of the local Institutional Animal Care and Use Committee.

**Cloning lentiCas9-Venus.** Venus template<sup>59</sup> was PCR amplified to add BamHI (5') and EcoRI (3') restriction sites (lowercase font) for cloning purposes using the following conditions: KOD buffer (1×), MgSO<sub>4</sub> (1.5 mM), dNTPs (0.2 mM each), forward primer (0.3 μM; [GGCCGGCCGgagtcGGCGCAACA AACTTCTCTCTGCTGAAACAAGCCGGAGATGTCTGAAGAGAATCCTGG ACCGATGGTGAGCAAGGGCGAGGA], reverse primer (0.3 μM; GGCCGGC CgaattcTTACTGTACAGCTCGTCCA), and KOD Hot Start DNA Polymerase (0.02 U μl<sup>-1</sup>) (Millipore). KOD PCR reaction used the following cycling conditions: 95 °C for 2 min; 50 cycles of 95 °C for 20 s, 60 °C for 20 s, and 70 °C for 30 s; 60 °C for 5 min. PCR products were purified (QIAquick PCR Purification Kit, Qiagen) and blunt end cloned with Zero Blunt PCR cloning kit (Invitrogen). PCR-blunt cloned products and lentiCas9-Blast (Addgene plasmid ID 52962) were separately digested with BamHI-HF (New England Biolabs) and EcoRI-HF (New England Biolabs) in 1× Buffer CutSmart at 37 °C (New England Biolabs). Digest of lentiCas9-Blast was performed to remove the blasticidin cassette. Then digested PCR product was ligated into the lentiCas9 backbone.

**Cloning lentiGuide-Crimson.** E2-Crimson template (Clontech) was PCR amplified to add BsiWI (5') and MluI (3') restriction sites for cloning purposes using the following conditions: KOD buffer (1×), MgSO<sub>4</sub> (1.5 mM), dNTPs (0.2 mM each), forward primer (0.3 μM; GGCCGGCCCGTACGctagcGCCACCATG GATAGCACTGAGAAGCTCATCAAGCCCTT), reverse primer (0.3 μM; GG CCGGCCacgcgtCTACTGGAACAGGTGGTGGCGGGCCT), and KOD Hot Start DNA Polymerase (0.02 U μl<sup>-1</sup>). KOD PCR reaction used the following cycling conditions: 95 °C for 2 min; 50 cycles of 95 °C for 20 s, 60 °C for 20 s, and 70 °C for 30 s; 60 °C for 5 min. PCR products were purified (QIAquick PCR Purification Kit) and cloned with Zero Blunt PCR cloning kit. Cloned products and lentiGuide-puro were separately digested with BsiWI (New England Biolabs) and MluI (New England Biolabs) in 1× buffer 3.1 at 37 °C (New England Biolabs). Digest of lentiGuide-Puro (Addgene plasmid ID 52963) was performed to remove the puromycin cassette. Then digested PCR product was ligated into the lentiGuide backbone.

**Cloning sgRNAs.** lentiGuide-Puro (Addgene plasmid ID 52963) was digested with BsmBI in 1× buffer 3.1 at 55 °C (New England Biolabs) for linearization. One unit of TSAP thermosensitive alkaline phosphatase (Promega) was added for 1 h at 37 °C to dephosphorylate the linearized lentiGuide and then TSAP was heat inactivated at 74 °C for 15 min. Linearized and dephosphorylated lentiGuide was run on an agarose gel and gel purified. sgRNA-specifying oligos were phosphorylated and annealed using the following conditions: sgRNA sequence oligo (10 μM); sgRNA sequence reverse complement oligo (10 μM); T4 ligation buffer (1×) (New England Biolabs); and T4 polynucleotide kinase (5 units) (New England Biolabs) with the following temperature conditions: 37 °C for 30 min; 95 °C for 5 min; and then ramp down to 25 °C at 5 °C min<sup>-1</sup>. Annealed oligos were ligated into lentiGuide in a 1:3 ratio (vector:insert) using T4 ligation buffer (1×) and T4 DNA Ligase (750 units) (New England Biolabs). Plasmids were verified by sequencing using a U6 promoter forward primer CGTAACTTG AAAGTATTTTCGATTCTTGGC.

sgRNA-specifying oligos using sgRNA sequences from the screen library (see Source Data associated with Figs 2 and 5) were obtained and cloned as described into either lentiGuide-Puro or lentiGuide-Crimson. sgRNA constructs were used to produce lentivirus and transduce HUDEP-2 with stable Cas9 expression. Bulk cultures were incubated for 7–10 days with 10 μg ml<sup>-1</sup> blasticidin and 1 μg ml<sup>-1</sup> puromycin selection to allow for editing. Then bulk cultures were plated clonally

at limiting dilution. Clones were allowed to grow for approximately 14 days and then genomic DNA was extracted using 50 μl QuickExtract DNA Extraction Solution per well.

**lentiTandemGuide cloning.** lentiGuide-sgRNA1 was digested with PspXI and XmaI at 37 °C for 4 h (New England Biolabs). Digests were run on an agarose gel and gel purified. lentiGuide-sgRNA2 was linearized using NotI (New England Biolabs). The hU6 promoter and sgRNA chimaeric backbone for lentiGuide-sgRNA2 was PCR amplified using the following conditions: KOD buffer (1×), MgSO<sub>4</sub> (1.5 mM), dNTPs (0.2 mM each), forward primer (0.3 μM; GGCCGGCC gctcgggGAGGGCCTATTTC), reverse primer (0.3 μM; CCGGCCGgcccggT TGTGGATGAATACTGCCATT), and KOD Hot Start DNA Polymerase (0.02 U μl<sup>-1</sup>) (Millipore). KOD PCR reaction used the following cycling conditions: 95 °C for 2 min; 50 cycles of 95 °C for 20 s, 60 °C for 20 s, and 70 °C for 30 s; 60 °C for 5 min. PCR products were purified (QIAquick PCR Purification Kit), blunt-ended cloned with Zero Blunt PCR cloning kit, transformed, and plated. Colonies were screened by digesting minipreps with EcoRI. Mini-preps were then digested with PspXI and XmaI as described above followed by PCR purification. After PCR purification, sgRNA2 was ligated into digested lentiGuide-sgRNA1. Sequence was verified with following primers: GGAGGCTTGGTAGGTTTA AGAA and CCAATTCCTACTCCTTTCAA.

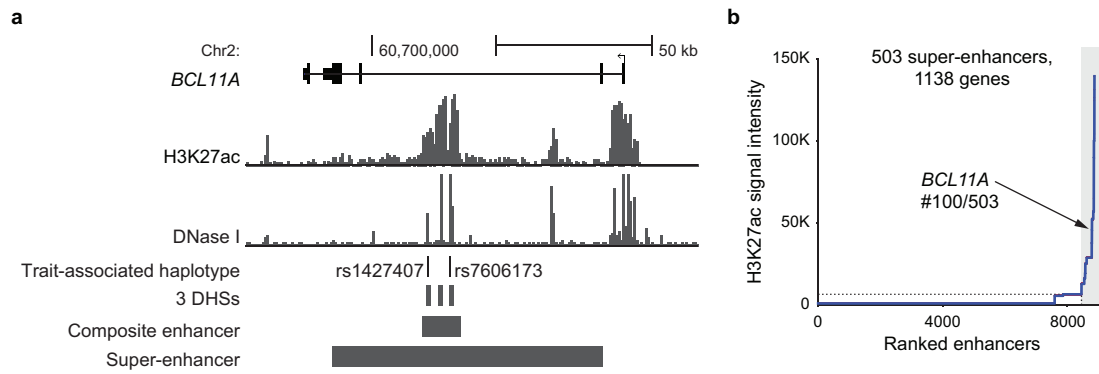
**Generation of HUDEP-2 with stable Cas9.** lentiCas9-Blast (Addgene plasmid ID 52962) or lentiCas9-Venus were produced as described above and used to transduce HUDEP-2 cells. Transduced cells were selected with 10 μg ml<sup>-1</sup> blasticidin or Venus<sup>+</sup> cells were sorted. Functional Cas9 was confirmed using the pXPR-011 (Addgene plasmid ID 59702) GFP reporter assay as previously described<sup>60</sup>.

**Generation of ey:mCherry reporter MEL cells.** A reporter MEL line in which mCherry was knocked into the *Hbb-y* locus was created (Extended Data Fig. 6c). Briefly, a TALEN-induced DSB was created adjacent to the *Hbb-y* transcriptional start site. A targeting vector with mCherry and a neomycin cassette was introduced through homology-directed repair. Homology arms included mm9 sequences from Chr7:111,001,667–111,002,675 and Chr7:111,000,661–111,001,666. Cre-mediated recombination was used to remove the neomycin cassette. Long-range PCR spanning each homology arm was used to ensure appropriate targeted integration. Cells were tested upon *Bcl11a* disruption by RT-qPCR and flow cytometry to confirm expected effects on ey:mCherry derepression. Subsequently, CRISPR-Cas9 was used as described above to produce cells with monoallelic composite enhancer deletion to maximize screening sensitivity for enhancer disruption.

**Generation of MEL cells with stable Cas9 expression.** lentiCas9-Blast (Addgene plasmid ID 52962) lentivirus was produced as described above and used to transduce MEL cells. Transduced cells were selected with 10 μg ml<sup>-1</sup> blasticidin. Functional Cas9 was confirmed using the pXPR-011 (Addgene plasmid ID 59702) GFP reporter assay as previously described<sup>60</sup>.

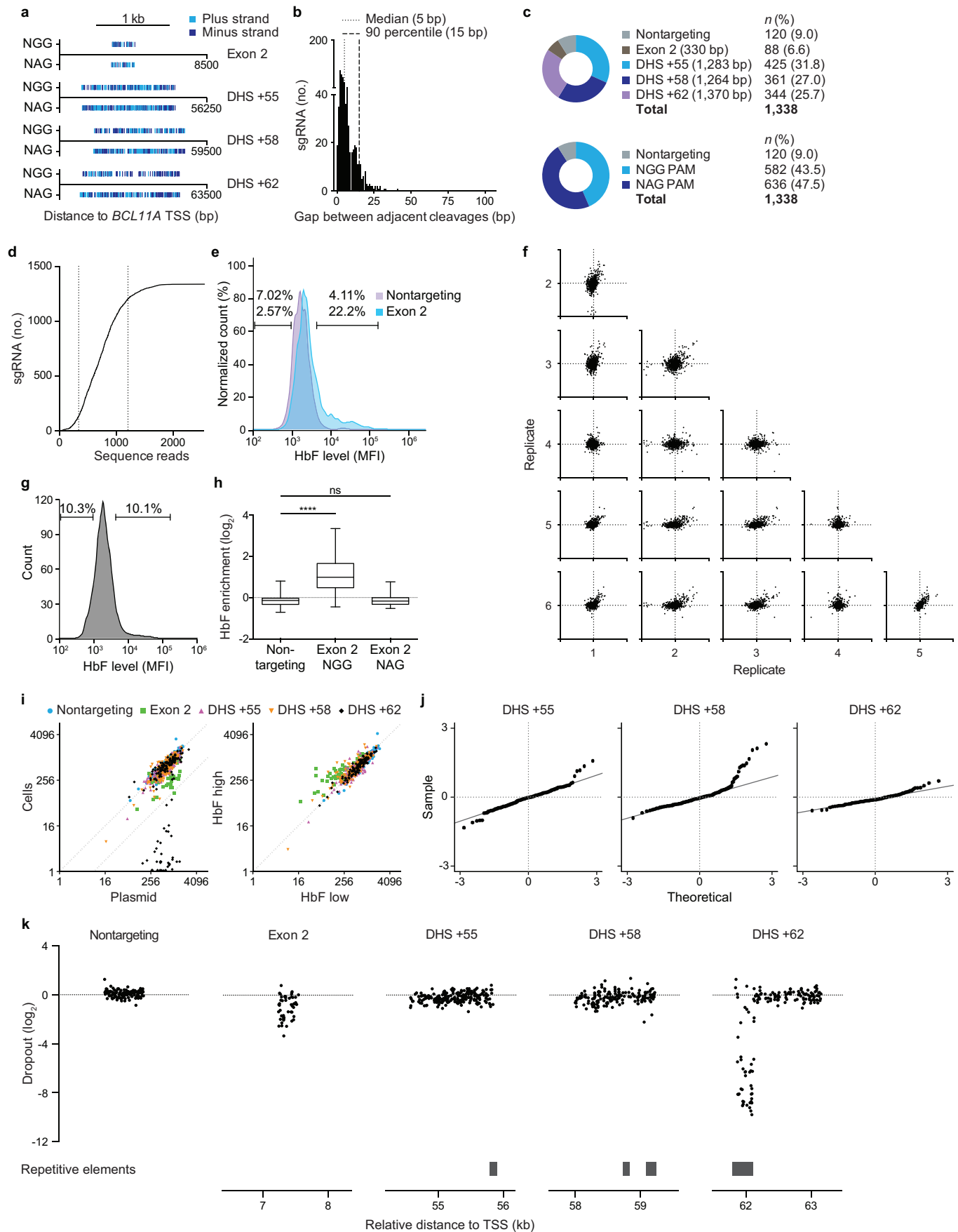
- Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods* **11**, 783–784 (2014).
- Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
- Giarratana, M. *et al.* Proof of principle for transfusion of *in vitro* generated red blood cells. *Blood* **118**, 5071–5079 (2011).
- Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
- Kowalczyk, M. S. *et al.* Intragenic enhancers act as alternative promoters. *Mol. Cell* **45**, 447–458 (2012).
- Dogan, N. *et al.* Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin* **8**, 16 (2015).
- Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
- Mathelier, A. *et al.* JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42**, 142–147 (2014).
- Weber, K., Bartsch, U., Stocking, C. & Fehse, B. A multicolor panel of novel lentiviral 'gene ontology' (LeGO) vectors for functional gene analysis. *Mol. Ther.* **16**, 698–706 (2008).
- Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnol.* **32**, 1262–1267 (2014).





**Extended Data Figure 1 | Human *BCL11A* locus.** **a**, Schematic of the human *BCL11A* locus (hg19, transcription from right to left) with erythroid chromatin marks and trait-associated haplotype denoted, and composite enhancer as

previously defined<sup>28</sup>. **b**, Ranked enhancers in primary human adult erythroid precursors by H3K27ac signal intensity, with super-enhancers shaded, and super-enhancer-associated genes indicated.

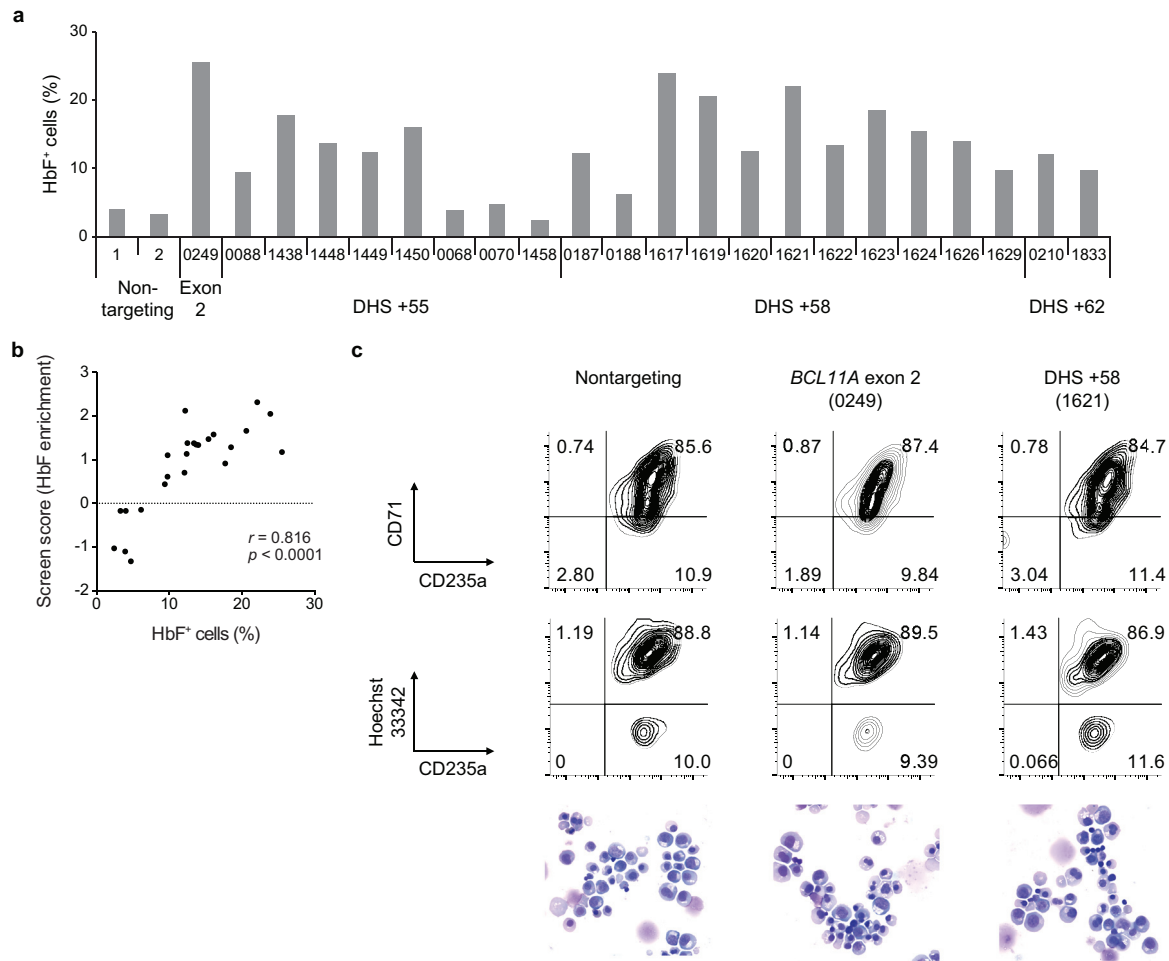


**Extended Data Figure 2 | Tiled pooled *in situ* CRISPR-Cas9 *BCL11A***

**enhancer screen.** **a**, Distribution of NGG and NAG PAM sgRNAs mapped to genomic cleavage position. The vertical lines represent cleavage sites for sgRNAs mapped to plus and minus strands. **b**, Gap distance between adjacent genomic cleavage position for NAG PAM sgRNAs. **c**, Library composition by target sequence and PAM restriction. **d**, Representation of both NGG and NAG sgRNA (1,338 sgRNAs in total) within the plasmid pool by deep sequencing. The median was 718 normalized reads and the 10th and 90th percentiles (indicated by the vertical dotted lines) ranged from 337 to 1,205 normalized reads. **e**, HbF distribution in HUDEP-2 cells transduced with Cas9 and individual sgRNAs, either non-targeting or targeting *BCL11A* exon 2.

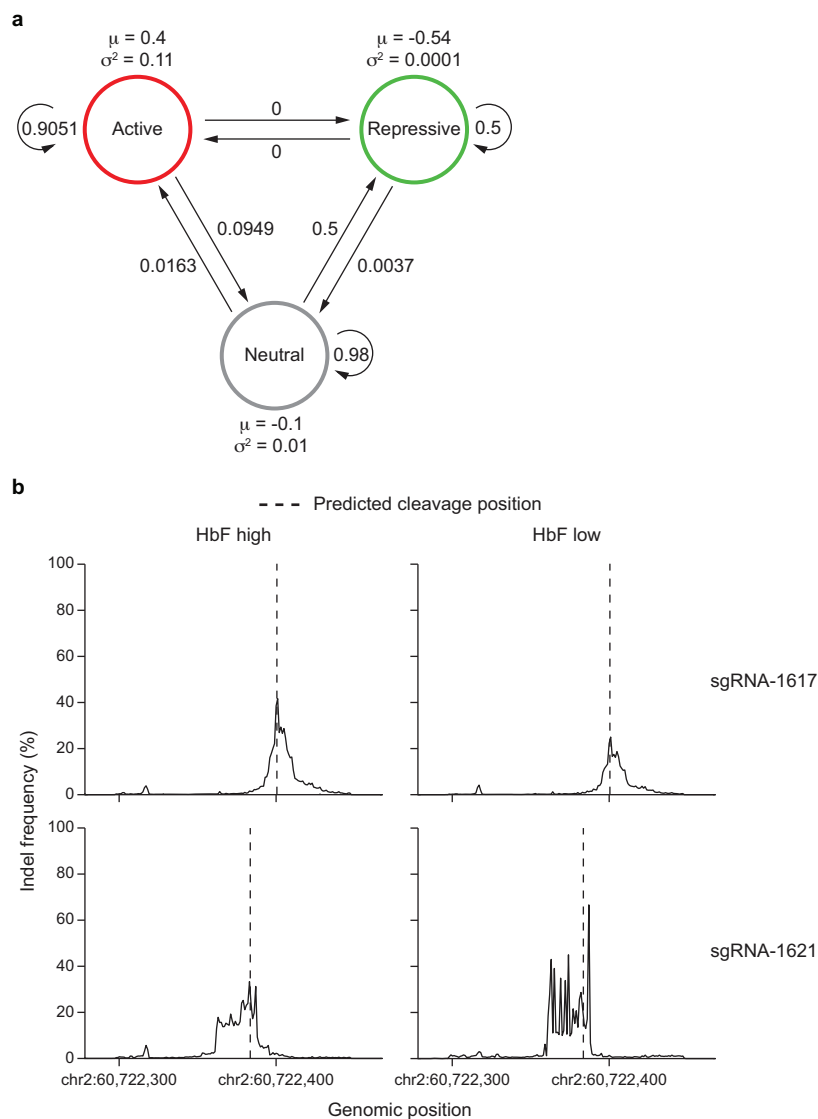
**f**, HbF enrichment scores of NGG sgRNAs in six biological replicates. **g**, Sort of library-transduced cells into HbF-high and HbF-low pools. **h**, Control sgRNA enrichment. Boxes demonstrate 25th, median, and 75th percentiles and whiskers minimum and maximum values. \*\*\*\* $P < 0.0001$ , NS, non-significant. **i**, NGG sgRNA representation in plasmid pool and cells at conclusion of experiment (left), and in HbF-high and HbF-low pools (right), with dotted lines at  $x = y$  and  $x = 8y$ . **j**, Quantile-quantile plots of NGG sgRNA enrichment scores. **k**, Cellular dropout scores of NGG sgRNAs relative to genomic cleavage position and repetitive elements. Non-targeting sgRNAs pseudo-mapped with 5-bp spacing.





**Extended Data Figure 3 | Validation of enhancer screen.** **a**, HbF<sup>+</sup> fraction in HUDEP-2 cells transduced in arrayed format with 24 sgRNAs from all 5 mapping categories with enrichment scores ranging from the highest to the lowest in the screen. **b**, Correlation between HbF enrichment score from pooled sgRNA screen and HbF<sup>+</sup> fraction by arrayed validation of individual

sgRNAs in HUDEP-2 cells. **c**, Erythroid differentiation of primary human erythroid precursors evaluated by CD71 and CD235a surface markers, enucleation frequency (CD235a<sup>+</sup> Hoescht33342<sup>-</sup>), and morphology by May-Grünwald Giemsa staining.

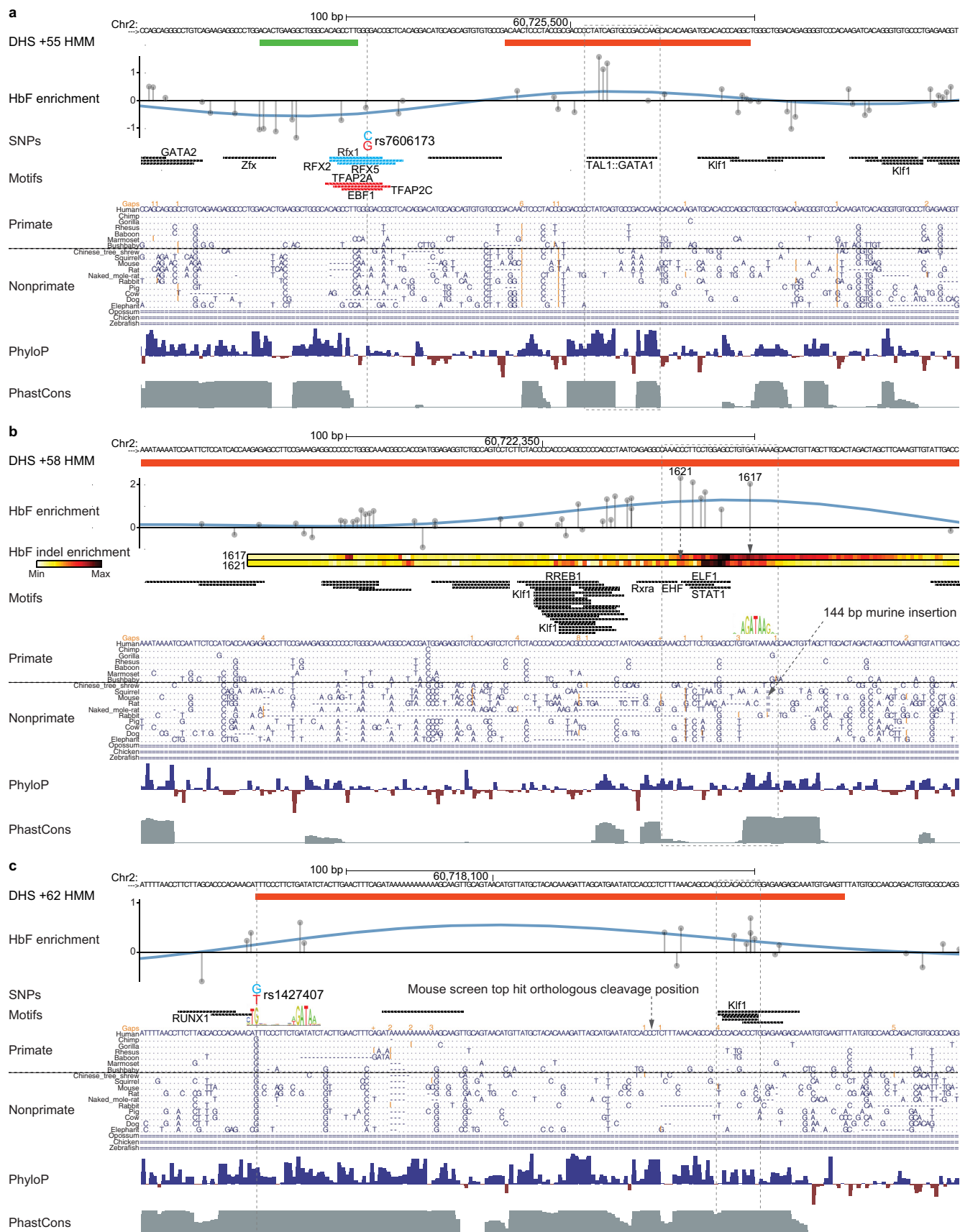


#### Extended Data Figure 4 | Functional assessment of enhancer sequences.

**a**, Topology of the HMM used to infer the three functional enhancer states (active, repressive and neutral). The emission probabilities of HbF enrichment scores from each state were modelled as Gaussian distributions (the values of  $\mu$  and  $\sigma^2$  are shown). The transition probabilities (arrows) are displayed.

**b**, Frequency distribution of indels from HUDEP-2 cells exposed to Cas9 and

individual sgRNAs, sorted into HbF-high and HbF-low pools, and subjected to deep sequencing of the target site. Indels calculated on a per nucleotide basis throughout an amplicon surrounding the sgRNA-1617 and -1621 cleavage sites (dotted lines). An indel enrichment ratio was calculated by dividing normalized indel frequencies in the HbF-high pool by those in the HbF-low pool.

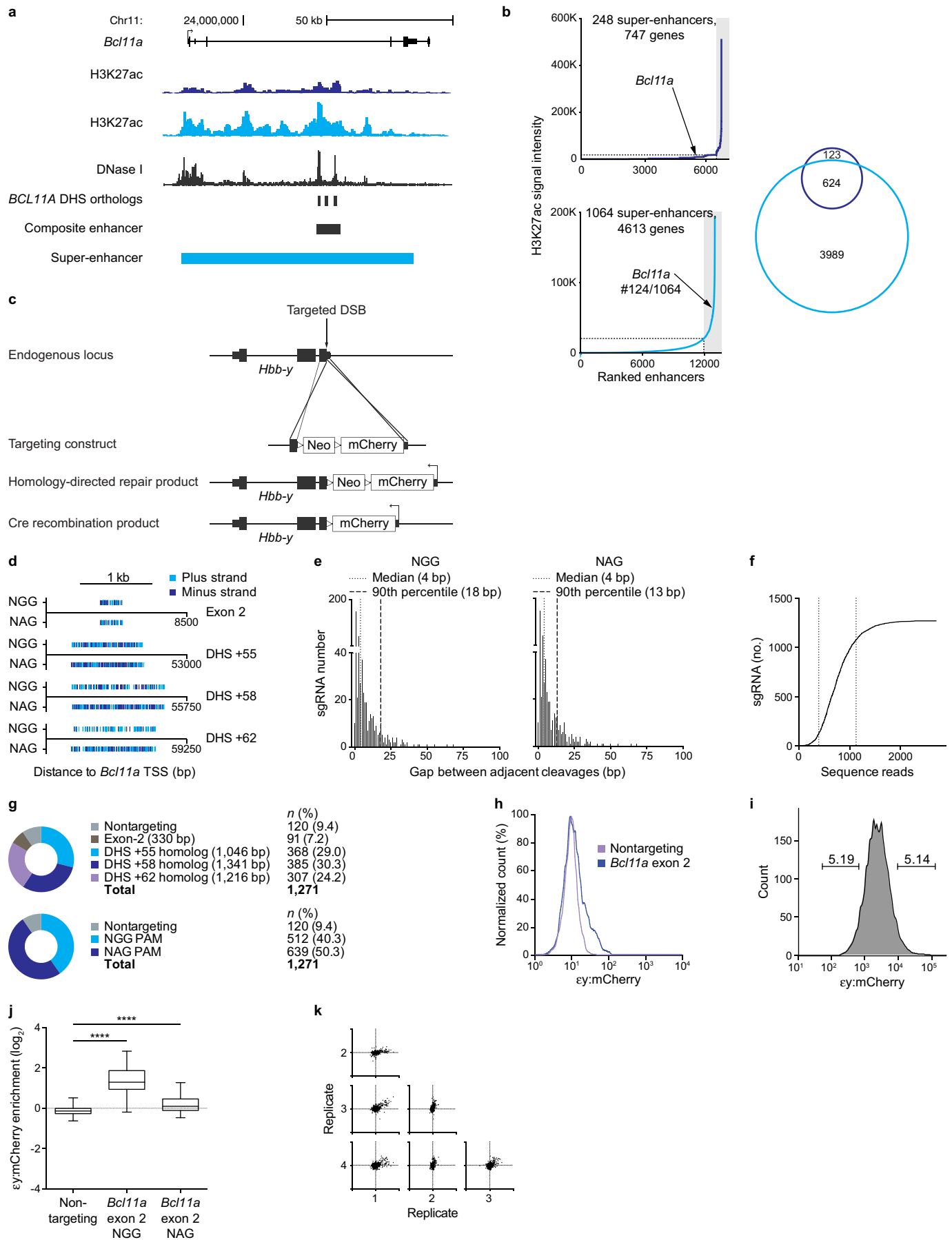




**Extended Data Figure 5 | Functional cores of the *BCL11A* enhancer.**

**a–c**, 200 bp at the functional cores of DHSs h+55, h+58 and h+62 defined by HMM states (active, red; repressive, green). HbF enrichment scores are shown by grey lines and circles. HbF indel enrichment per nucleotide based on amplicon genomic sequencing of sorted cells exposed to either sgRNA-1617 (top) or -1621 (bottom) is shown. Common SNPs (MAF > 1%) are shown with dotted lines with HbF-low allele in blue and HbF-high allele in red; no common SNPs are present at the h+58 region. JASPAR motifs ( $P < 10^{-4}$ ) are depicted in black except for those with allele-specific significance depicted by allelic colour. Selected motifs annotated by transcription factor on the basis

of known erythroid-specific function or genomic position. Motif LOGOs at key positions with motif scores  $P < 10^{-3}$  as described in text. Dotted boxes show regions of highest HbF enrichment score at each core with underlying predicted motifs. Orthologous sequences are listed from representative primates and nonprimates of distributed phylogeny. PhyloP (scale from -4.5 to 4.88) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 100 vertebrates are shown. An arrow indicates a 144 bp insertion in the mouse genome relative to the human reference adjacent to the orthologous GATA1 motif at h+58.

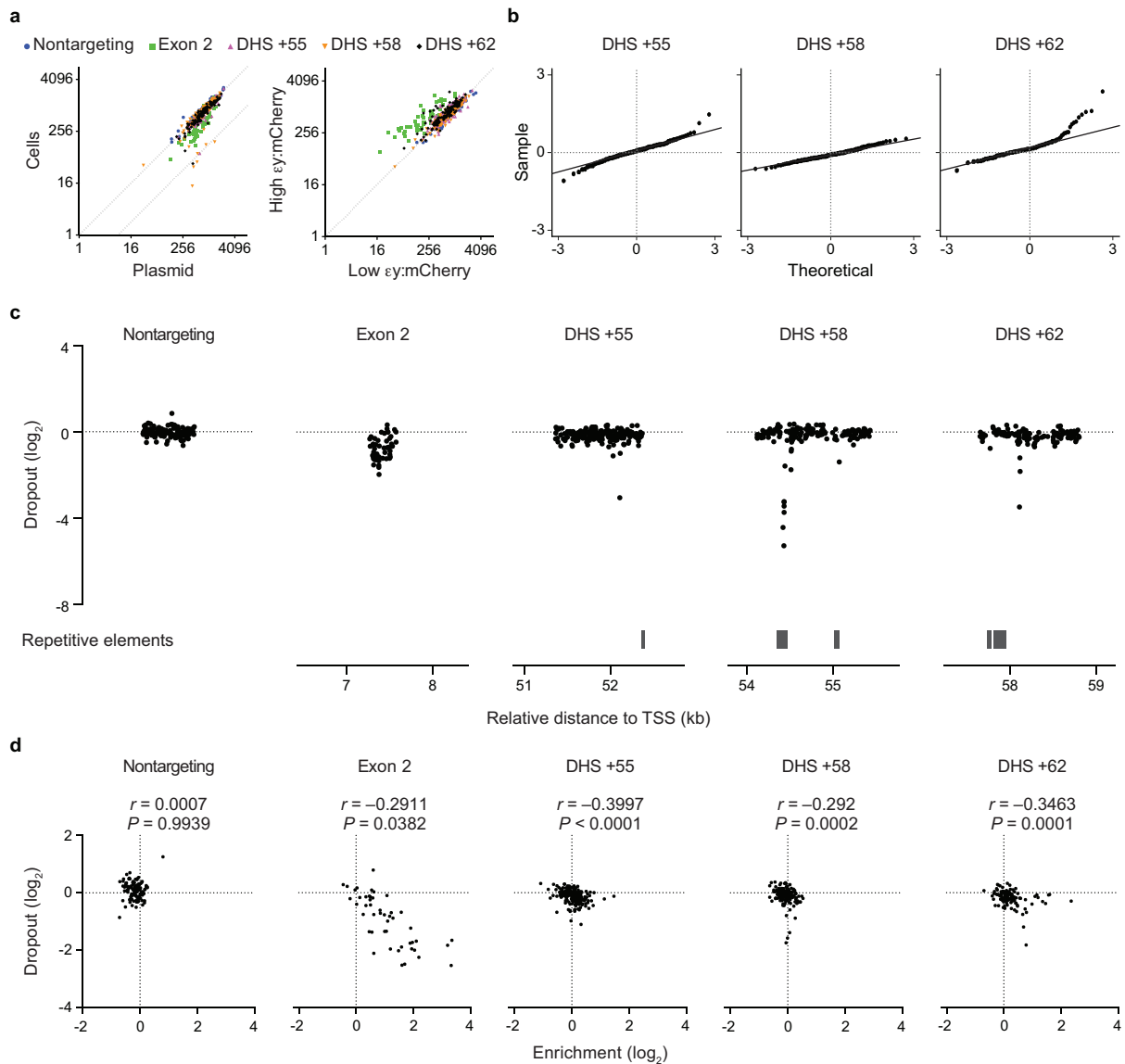


**Extended Data Figure 6 | Tiled pooled *in situ* CRISPR-Cas9 *Bcl11a***

**enhancer screen.** **a**, Schematic of the mouse *Bcl11a* locus (mm9, transcription from left to right) with erythroid chromatin marks (top, dark blue H3K27ac from ref. 55; middle, light blue H3K27ac from ref. 56; and bottom, black DNase I from ref. 28) and regions of primary sequence homology to the human DHSs displayed. *y* axes for H3K27ac tracks are both scaled to maximum 3.5 reads per million. Composite enhancer as previously defined<sup>28</sup>. **b**, Ranked enhancers in mouse erythroid precursors by H3K27ac signal intensity<sup>55,56</sup>, with super-enhancers shaded. Super-enhancer associated genes are indicated by Venn diagram. **c**, Strategy to knock-in by homology-directed repair the fluorescent protein mCherry into the mouse embryonic globin *Hbb-y* locus (encoding the  $\epsilon\gamma$  embryonic globin chain). **d**, Distribution of NGG and NAG PAM sgRNAs mapped to genomic cleavage position with vertical lines

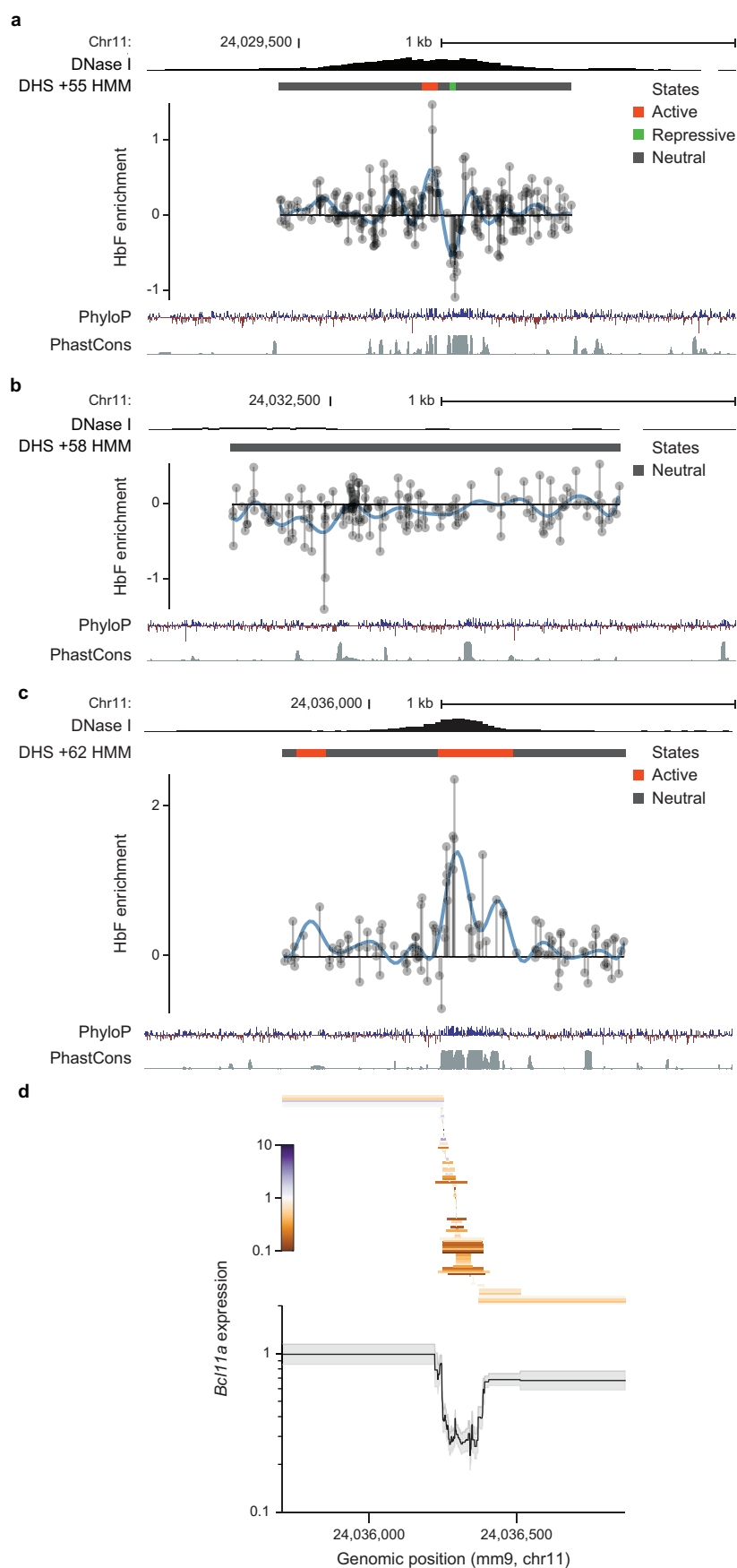
representing cleavage sites for sgRNAs mapped to plus and minus strands. **e**, Distance to adjacent genomic cleavage position for NGG (left) and NAG (right) PAM sgRNAs. **f**, Representation of the 1,271 NGG and NAG sgRNAs within the plasmid pool by deep sequencing. The median was 735 normalized reads and the 10th and 90th percentiles (indicated by the vertical dotted lines) ranged from 393 to 1,240 normalized reads. **g**, Library composition by target sequence and PAM restriction. **h**, mCherry expression upon exposure to Cas9 and an individual NGG sgRNA targeting *Bcl11a* exon 2 in MEL  $\epsilon\gamma$ :mCherry reporter cells. **i**,  $\epsilon\gamma$ :mCherry sort of library transduced cells. **j**, Control sgRNA enrichment. Boxes demonstrate 25th, median and 75th percentiles and whiskers minimum and maximum values. \*\*\*\* $P < 0.0001$ . **k**, Enrichment scores of NGG sgRNAs between four biological replicates.





**Extended Data Figure 7 | *Bcl11a* enhancer screen analyses.** **a**, NGG sgRNA representation in plasmid pool and cells at conclusion of experiment (left), and in  $\epsilon y:mCherry$ -high and  $\epsilon y:mCherry$ -low pools (right), with dotted lines at  $x = y$  and  $x = 8y$ . **b**, Quantile-quantile plots of NGG sgRNA  $\epsilon y$  enrichment

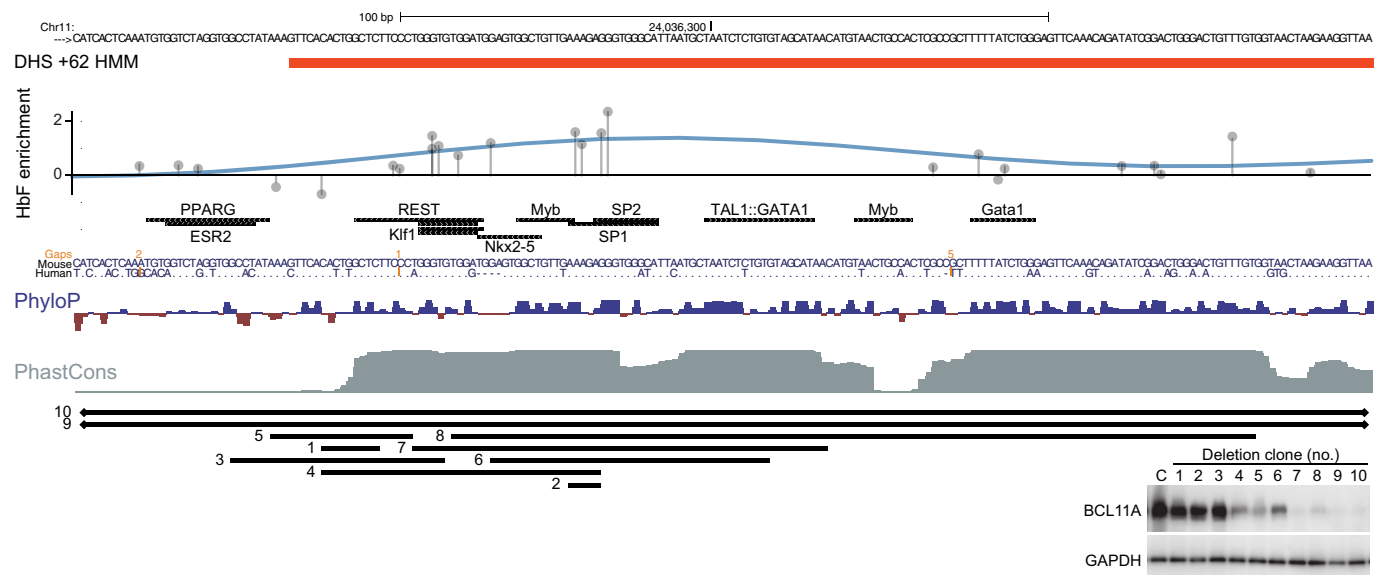
scores. **c**, Cellular dropout scores of NGG sgRNAs relative to genomic cleavage position and repetitive elements. Non-targeting sgRNAs are pseudo-mapped with 5 bp spacing. **d**, Correlation between cellular dropout and  $\epsilon y$  enrichment scores.



**Extended Data Figure 8 | Functional sequences at the *Bcl11a* erythroid enhancer.** **a–c**, HMM segmentation of active functional states at m+55, m+58 and m+62 orthologues. HbF enrichment scores are shown as grey lines and circles with the blue line representing smoothed enrichment score. DNase I sequencing from mouse fetal liver erythroid precursors<sup>28</sup>. PhyloP (scale from –3.3 to 2.1) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 30 vertebrates are shown. **d**, Top: *Bcl11a* expression determined by RT–qPCR displayed as a heat map in 108 hemizygous m+62

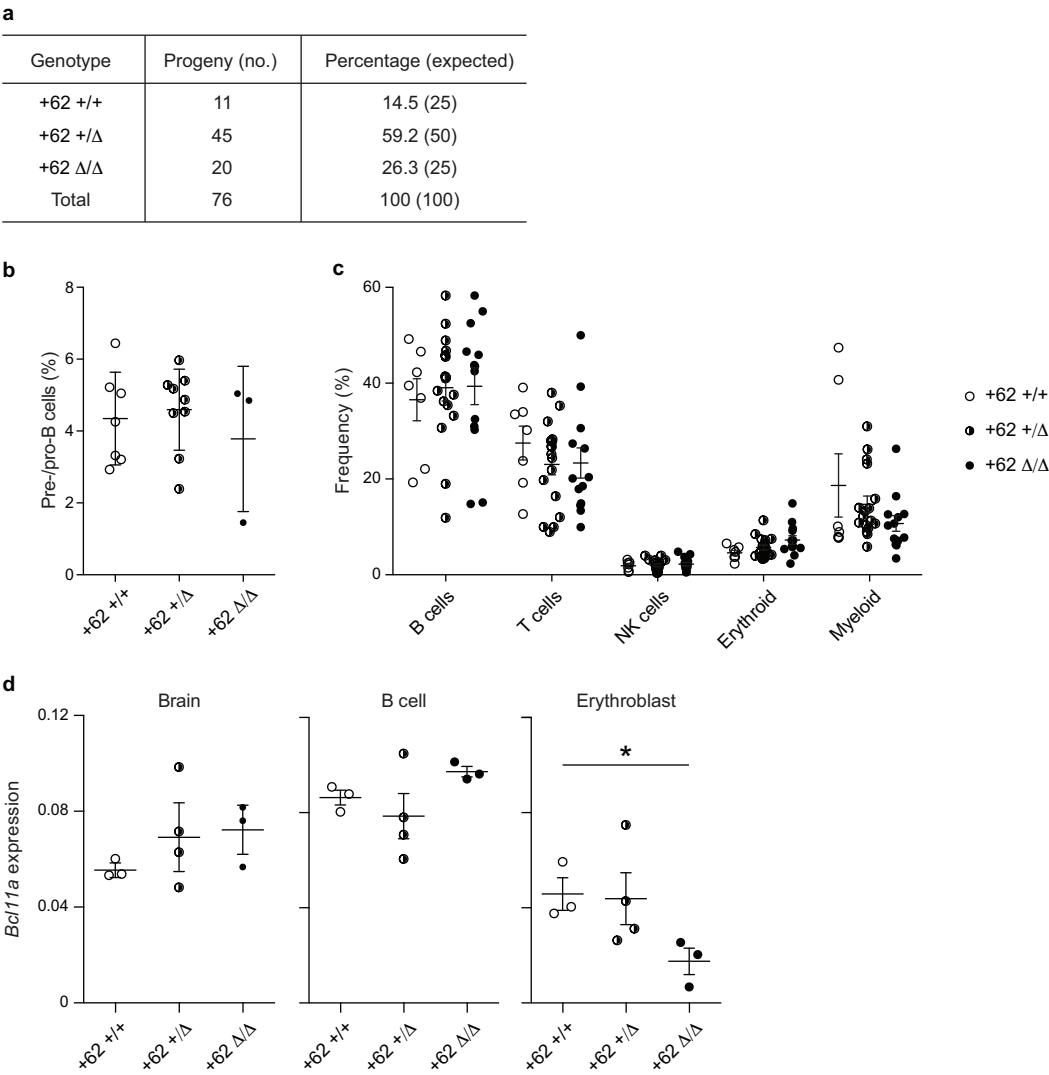
orthologue deletion clones ordered by genomic position of deletion midpoint. Each bar demonstrates the genomic position of the deletion breakpoints and the associated colour demonstrates the level of *Bcl11a* expression. Bottom: *Bcl11a* expression determined by RT–qPCR in 108 hemizygous m+62 orthologue deletion clones. Per nucleotide mean effect size was calculated as the mean fold change in *Bcl11a* expression from all clones in which that nucleotide was deleted. Grey shading represents one s.d. The *Bcl11a* expression data are shown with same *x* axis as in panel c immediately above.





**Extended Data Figure 9 | Evaluation of the m+62 functional core.** 200 bp at the functional core of the m+62 orthologue defined by HMM state. Enrichment scores are shown as grey lines and circles with the blue line representing smoothed enrichment score. JASPAR motifs ( $P < 10^{-4}$ ) are depicted with selected motifs annotated by transcription factor name on the basis of known erythroid-specific function or genomic position. Orthologous

human sequences are listed. PhyloP (scale from  $-3.3$  to  $2.1$ ) and PhastCons (from  $0$  to  $1$ ) estimates of evolutionary conservation among 30 vertebrates are shown. Individual numbered hemizygous deletion clones with indicated breakpoints were evaluated by BCL11A immunoblot (C, control). Clones 9 and 10 encompass the entire m+62 orthologue.



**Extended Data Figure 10 | Requirement of *Bcl11a* erythroid enhancer during murine ontogeny.** **a**, Progeny of heterozygous *Bcl11a* m+62 orthologue deletion intercrossores as compared to expected Mendelian ratio. **b**, Fraction of fetal liver comprised of B-cell progenitors at E16.5 from various genotypes. **c**, Peripheral blood analysis from 4-week-old mice to examine the

frequency of various circulating haematopoietic lineages in *Bcl11a* m+62 orthologue deletion wild-type, heterozygous, and homozygous mice. **d**, *Bcl11a* expression in β-YAC/+62 deletion mice (each symbol represents the mean expression from technical replicates from an individual mouse). \**P* < 0.05; error bars represent s.e.m.

# Cryo-electron microscopy structure of the Slo2.2 Na<sup>+</sup>-activated K<sup>+</sup> channel

Richard K. Hite<sup>1</sup>, Peng Yuan<sup>1</sup>, Zongli Li<sup>2</sup>, Yichun Hsuing<sup>1</sup>, Thomas Walz<sup>2</sup> & Roderick MacKinnon<sup>1</sup>

Na<sup>+</sup>-activated K<sup>+</sup> channels are members of the Slo family of large conductance K<sup>+</sup> channels that are widely expressed in the brain, where their opening regulates neuronal excitability. These channels fulfil a number of biological roles and have intriguing biophysical properties, including conductance levels that are ten times those of most other K<sup>+</sup> channels and gating sensitivity to intracellular Na<sup>+</sup>. Here we present the structure of a complete Na<sup>+</sup>-activated K<sup>+</sup> channel, chicken Slo2.2, in the Na<sup>+</sup>-free state, determined by cryo-electron microscopy at a nominal resolution of 4.5 Å. The channel is composed of a large cytoplasmic gating ring, in which resides the Na<sup>+</sup>-binding site and a transmembrane domain that closely resembles voltage-gated K<sup>+</sup> channels. In the structure, the cytoplasmic domain adopts a closed conformation and the ion conduction pore is also closed. The structure reveals features that can explain the unusually high conductance of Slo channels and how contraction of the cytoplasmic gating ring closes the pore.

Potassium channels control the excitability of electrically active cells by regulating the resting membrane potential in response to a variety of stimuli<sup>1</sup>. One such stimulus is an increase in intracellular Na<sup>+</sup>, which occurs following repeated membrane depolarization. Slo2.2, also known as Slack or KCNT1, is opened by increases in intracellular Na<sup>+</sup> (refs 2–8). This channel is widely expressed in the brain and shapes neuronal excitability, especially in neurons that fire action potentials at high frequencies<sup>8–12</sup>. Mutations in *KCNT1*, the gene that encodes the Slo2.2 protein, are linked to a variety of intellectual disabilities including malignant migrating partial seizures in infancy<sup>13–17</sup>, autosomal dominant nocturnal frontal lobe epilepsy<sup>18</sup> and Ohtahara syndrome<sup>19</sup>. Slo2.2 has also been found in other cell types, including nociceptive and sensory neurons, where it is hypothesized to influence pain sensitivity, and in epithelial cells of the thick ascending limb of Henle's loop, where it is involved in ion reabsorption<sup>20–23</sup>.

Slo2.2 is a member of the Slo family of large conductance K<sup>+</sup> channels, which are characterized by a transmembrane domain (TMD) containing six or seven transmembrane helices, and a large cytoplasmic domain (CTD) containing two regulator-of-K<sup>+</sup>-conductance (RCK) domains. High-resolution structural data do not currently exist for a full-length Slo channel, but structures of isolated CTDs of Slo1, a Ca<sup>2+</sup>- and voltage-activated channel, were determined in Ca<sup>2+</sup>-free (closed) and Ca<sup>2+</sup>-bound (open) conformations<sup>24–26</sup>. In these structures the CTDs are organized into tetrameric gating rings that expand upon Ca<sup>2+</sup> binding<sup>25</sup>. This expansion appears to be sufficient to open an inner helix gate in the transmembrane channel<sup>25</sup>. However, solvent accessibility experiments carried out in Slo1 and in Slo2.1 (a channel that shares a high degree of sequence similarity to Slo2.2, also Na<sup>+</sup>-activated) have led to a hypothesis that these channels never fully close an inner gate, but instead close at the selectivity filter<sup>27–29</sup>. Thus, the basic question of where the gate is in Slo channels has remained unanswered.

## Cryo-electron microscopy analysis

To determine the structure of a full-length Slo2.2 channel, a construct encoding the entire chicken *KCNT1* gene (84% sequence identity to human *KCNT1*; Extended Data Fig. 1) was heterologously expressed

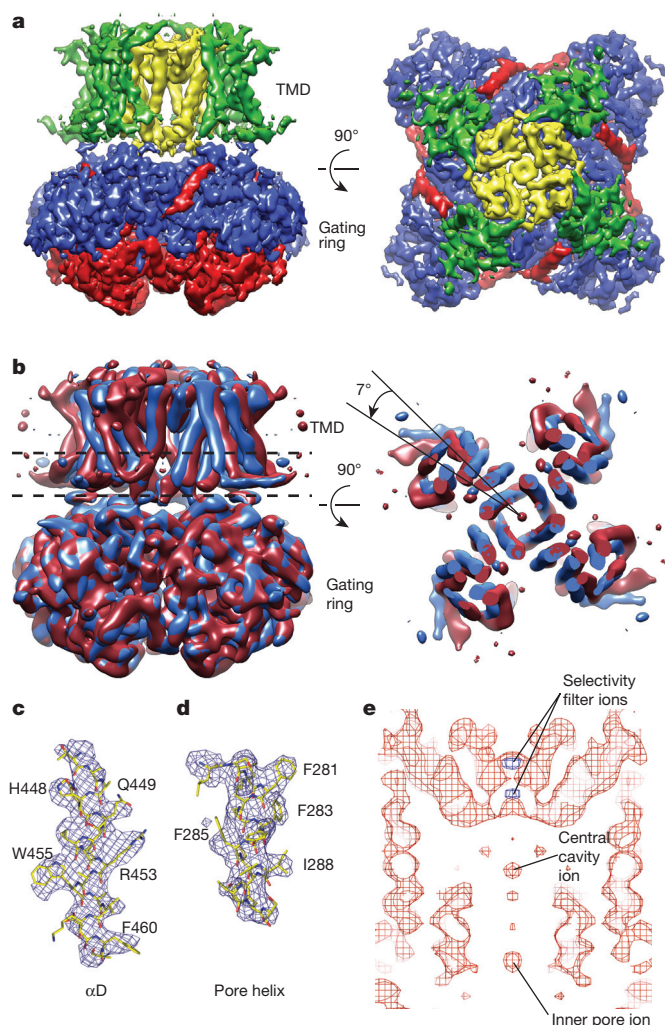
in *Spodoptera frugiperda* cells. Images of frozen-hydrated preparations of detergent- and lipid-solubilized Slo2.2 tetramers in the absence of Na<sup>+</sup> were recorded using a direct electron detector (Extended Data Fig. 2a). Using single-particle analysis, a cryo-electron microscopy (cryo-EM) density map of a Slo2.2 tetramer was calculated at a nominal resolution of 4.5 Å with C<sub>4</sub> symmetry imposed (Extended Data Fig. 2d and Extended Data Table 1). The density map contains two domains into which the tetrameric structures of a Slo1 gating ring<sup>24</sup> and the voltage-dependent K<sup>+</sup> channel (Kv) chimera α-subunit<sup>30</sup> could be manually fitted.

During fitting it became apparent that the density corresponding to the periphery of the TMD was of poorer quality than the density corresponding to the gating ring (Extended Data Fig. 3a). To assess the source of disorder in the TMD, 3D maximum-likelihood classification of the particle images was performed, yielding five similar but non-identical subclasses. These subclasses were related by a rotation of the TMD with respect to the gating ring about the fourfold axis (Fig. 1b). In the two most extreme subclasses, the rotation angle is 7°. Consequently, the larger mass of the gating rings resulted in them being well aligned in the reconstruction, whereas the smaller TMD is blurred, especially at the perimeter furthest from the fourfold axis. Separate focused refinements of the TMD and the gating ring using soft masks improved the maps for both domains (Fig. 1a) and provided enough detail to build a model starting with a voltage-dependent K<sup>+</sup>-channel structure for the TMD and a Slo1 gating-ring structure for the gating ring (Fig. 1c–e, Extended Data Fig. 6 and Extended Data Table 1). Strong K<sup>+</sup> ion density (6.5 σ) is present in the selectivity filter and weaker density (4 σ) is present in the central cavity and in the inner pore closer to the cytoplasm (Fig. 1e and Extended Data Fig. 5a). Prior knowledge of the location of K<sup>+</sup> ions in the selectivity filter and central cavity of K<sup>+</sup> channels from X-ray crystallographic studies<sup>31</sup> supports our assignment of these densities as ions rather than noise along the fourfold axis (Extended Data Fig. 5b).

The gating-ring and TMD models were built and refined independently against phases and amplitudes calculated from the focused refinement maps using reciprocal-space algorithms. The gating ring

<sup>1</sup>Rockefeller University and Howard Hughes Medical Institute, 1230 York Avenue, New York, New York 10065, USA. <sup>2</sup>Department of Cell Biology and Howard Hughes Medical Institute, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.





**Figure 1 | Cryo-EM structure of chicken Slo2.2.** **a**, Cryo-EM density map of chicken Slo2.2 following focused refinements of the TMD and the gating ring. The gating-ring map is filtered to 4.2 Å and the TMD map is filtered to 4.5 Å. The S1–S4 domain is green, the pore domain is yellow, the RCK1 domain is blue and the RCK2 domain is red. **b**, Density map of two of the 3D subclasses (coloured in red and blue) filtered to 7 Å and aligned by their gating rings. The density slice corresponds to the region of the TMD between the dashed lines. **c**, **d**, Fragments of the density map corresponding to helix  $\alpha$ D in the gating ring (**c**) and the pore helix in the TMD (**d**). The refined model is shown in stick representation. Large side chains that were used to register the sequence are labelled. **e**, Central section of the density map in **a** through the TMD calculated at 4  $\sigma$  (red) and 6.5  $\sigma$  (blue) with densities corresponding to K<sup>+</sup> ions labelled.

was refined to a resolution of 4.2 Å using *P*<sub>4</sub> symmetry, the nominal resolution of the map, with the ratio of the residual factors  $R_{\text{work}}/R_{\text{free}} = 25.5/28.0$  and good geometry (Extended Data Figs 3d, 4a, b and Extended Data Table 1). The nominal resolution of the TMD map is 5.2 Å, but local resolution estimates combined with the resolved features of the map indicate that the map, particularly within the pore domain, contains information that extends to higher resolution (Fig. 1a, c–e, Extended Data Fig. 3c, d and Extended Data Table 1). The TMD structure was therefore refined using *P*<sub>4</sub> symmetry to 4.5 Å, the nominal resolution of the full channel map, with  $R_{\text{work}}/R_{\text{free}} = 25.7/27.1$  and good geometry (Extended Data Fig. 4a, c and Extended Data Table 1). A model of the complete channel was generated by manually docking the independently refined TMD and gating-ring models into the full channel map. The fit was improved by rigid-body refinement of the gating ring and the TMD against

phases and amplitudes extracted from the full map (Fig. 2b and Extended Data Table 1). The final refined model contains residues 244–337, 351–612, 769–860, 865–1019, 1041–1171. In addition, unregistered polyaniline  $\alpha$ -helical models were built for S0 (amphipathic helix similar to those present in Kv channels), S1, the linker between S1 and S2, S2, S3 and S4 in the TMD, and  $\alpha$ X',  $\alpha$ Y' and  $\alpha$ Z' in the gating ring (Extended Data Fig. 1).

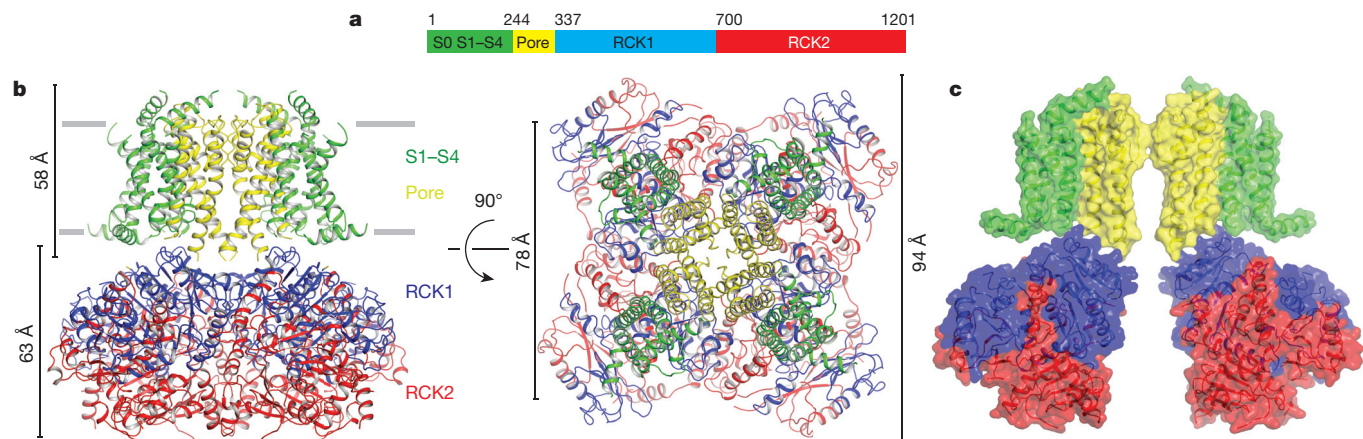
## Structure of Slo2.2

The tetrameric Slo2.2 channel is 120 Å in length along an axis perpendicular to the membrane plane (Fig. 2b). More than half of this length corresponds to the gating ring, which protrudes into the cytoplasm. Viewed down the fourfold axis, the TMD is approximately 80 Å on the square edge and the gating ring is 95 Å (Fig. 2b). Each subunit polypeptide chain builds up the Slo2.2 channel in three layers: a TMD layer and two RCK domain layers (RCK1 and RCK2), which form the gating ring (Fig. 2b, c).

The TMD is similar in topology to a Kv channel, to which Slo2.2 is related<sup>32</sup>. The central ion conduction pore is surrounded in the membrane by voltage sensor-like domains, which we refer to as S1–S4 domains to emphasize their voltage-independent uniqueness (Fig. 2b, c). As in Kv channels, the S1–S4 domains are arranged as appendages that project laterally into the lipid membrane. At a level of greater structural detail, we note two important distinctions between the TMDs of Kv1.2 and Slo2.2.

The first distinction is the manner in which the surface of the S1–S4 domain forms its contacts with the pore. In the Kv channel, the voltage sensor's interaction with the pore occurs mainly through the extracellular half of S1 and through the S4–S5 linker (Fig. 3). This configuration leaves the charged S4 helix free to move in response to voltage differences across the membrane and exert force on the S4–S5 linker to gate the channel. In Slo2.2, the extracellular half of S1, nearly the entire S4 and a helix from the S1–S2 connecting segment form an extensive contact surface with the pore. The Kv voltage sensor and Slo2.2 S1–S4 domains have essentially the same helical structure, but their different orientations with respect to the pore reconfigure the contact surface (Fig. 3). In Slo2.2, S4 runs antiparallel to and in close contact with S5, as if its function is to help glue the S1–S4 domain to the pore rather than to provide a mobile voltage sensor. This possibility is consistent with the amino acid sequence of Slo2.2, which shows two positive- and two negative-charged amino acids in or near S4 instead of the excess positive charges observed in the S4 of a Kv channel and in the S4 sequence of Slo1 and Slo3 (ref. 33). Furthermore, mutations involving these positive and negative charges in Slo2.2 do not produce large changes in voltage-dependence<sup>34</sup>. Thus, S1–S4 does not appear to function as a mobile voltage sensor in Slo2.2. A surprising observation, which may be related to the differences in S4, is that we cannot detect cryo-EM density for an S4–S5 helical linker in Slo2.2. Higher-resolution data will be required to tell whether this linker is partially disordered or simply not present as a helix. Its absence in Slo2.2 is conspicuous, and implies that this region serves a different function than in Kv channels.

The second distinction is the manner in which the TMD connects to the gating ring. The Kv channel contains a CTD consisting of a T1 domain and  $\beta$ -subunit. These are held at a distance of nearly 15 Å away from the membrane plane and consequently ions can gain access to the transmembrane pore through lateral openings between the TMD and the CTD, near the membrane surface<sup>35</sup>. In contrast, in Slo2.2 the TMD and gating ring form a complementary interface near the membrane surface. This is most easily appreciated in a surface representation of Slo2.2 (Fig. 2c): the gap at the perimeter of the interface would be filled in if a full side-chain model (instead of polyaniline) of the S1–S4 domain had been built. The relatively snug fit between the TMD and gating ring in Slo2.2 has important implications for ion conduction and gating.



**Figure 2 | Architecture of Slo2.2.** **a**, Domain organization of chicken Slo2.2. **b**, Ribbon diagram of Slo2.2. The S1–S4 domain is green, the pore domain is yellow, the RCK1 domain is blue and the RCK2 domain is red. The

approximate width of the membrane is marked by the grey lines. **c**, Surface depiction of Slo2.2 with front and rear subunits removed for clarity.

### Ion conduction

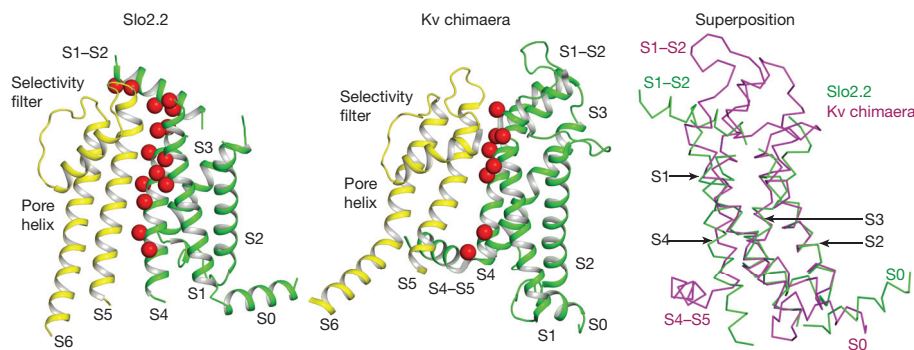
In contrast to the Kv channel, the continuous pore across the membrane and through the gating ring makes Slo2.2 more like an inward-rectifier  $K^+$  channel, which also has an extended ion pathway, owing to a tightly engaged CTD<sup>36,37</sup> (Fig. 4a). In inward rectifiers, however, the extended pore is narrower and contains binding sites for blocking ions, a property that gives rise to inward rectification<sup>36,38</sup>. In contrast, in Slo2.2 the gating ring creates a massive funnel that begins nearly 40 Å wide and gradually narrows as it approaches the pore inside the TMD. The inner surface of the funnel is highly electronegative, owing to the presence of many aspartate and glutamate amino acids (Fig. 4c). This electrostatically negative funnel, by functioning as a cation attractor, undoubtedly contributes to the unusually high conductance (nearly 200 pS for Slo2.2: about 20 times that of a canonical Kv channel such as Shaker<sup>39</sup>) of Slo channels (Extended Data Fig. 7a). Indeed, Slo channels are the highest-conductance  $K^+$  channels known, and a recent mutagenesis study with Slo1 showed that the gating ring contributes to the high conductance<sup>40</sup>.

The ion conduction pathway leading from the cytoplasm up to the selectivity filter has a single constriction that is formed by the crossing of four transmembrane S6 (inner) helices (Fig. 4b). The pore constricts to a minimum radius at Met333, which points its  $\beta$ -carbon towards the centre of the pore. Cryo-EM density for this side chain beyond the  $\beta$ -carbon is not well resolved; however, modelling the six most frequently observed rotamers of Met yields a range of pore diameters—defined by diagonally opposed van der Waals surfaces—ranging from 4 Å to 6 Å (Fig. 4b and Extended Data Fig. 8).

A hydrated  $K^+$  ion has a diameter closer to 8 Å; therefore, the structure is compatible with a closed conformation, consistent with the  $Na^+$ -free conditions under which this structure was determined. Adding to the physical barrier that the van der Waals constriction at the level of Met333 would present to a permeating  $K^+$  ion, the local electrostatic environment would raise the barrier further (Fig. 4c). The band of blue colour seen in Fig. 4c (positive electrostatic potential) coincident with the constriction is due to Arg335 and Lys337 on the inner helices and Arg396 on the surface of the gating ring; these positive-charged amino acids impose a long-range cation-repulsive potential. Thus, the inner helix gate appears closed to  $K^+$  diffusion by virtue of its narrow constriction and positive electrostatic potential at the level of the constriction.

### Gating inferences from structure

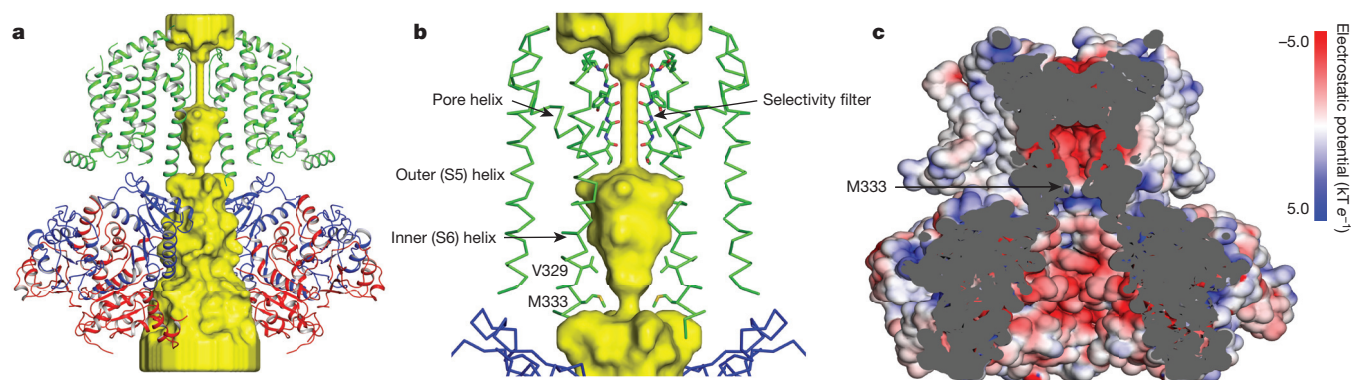
The Slo2.2 gating ring is similar in overall structure to the gating rings in Slo1 and Slo3 (refs 24, 25, 41). Two non-identical RCK domains on the C terminus of each subunit assemble through a 'flexible interface'; four RCK domain pairs then assemble with each other through 'assembly interfaces' to form the ring (Fig. 5). The assembly interface is essential to the function of Slo1 because it contains the high-affinity  $Ca^{2+}$ -binding site known as the ' $Ca^{2+}$  bowl'<sup>26</sup>. Instead of a  $Ca^{2+}$  bowl, Slo2.2 has an extended  $\alpha$ -helix,  $\alpha Q'$ , which interacts across the assembly interface with an adjacent subunit (Fig. 5). A helix–turn–helix structure ( $\alpha X'$ ,  $\alpha Y'$  and  $\alpha Z'$ ) is also unique to the assembly interface in Slo2.2. This helical elaboration is provided by an approximately 110-residue insertion in the RCK2 domain that is conserved and



**Figure 3 | Interactions between pore and S1–S4 domains.** Ribbon diagram of the S1–S4 domain of Slo2.2 (left) and the voltage-sensor domain of Kv chimera (middle). The S1–S4 and voltage-sensor domains are green and the pore domains are yellow. Red spheres represent the residues on S1, S4 and the

S1–S2 linker that are close enough to interact with the pore domain. Superposition (right) of the S1–S4 domain of Slo2.2 (green) with the voltage-sensor domain of Kv chimera (magenta) by aligning S1, S2 and S3.





**Figure 4 | Slo2.2 ion conduction pathway.** **a**, Surface representation of the Slo2.2 pore and ribbon diagram with front and rear subunits excluded for clarity. The  $K^+$ -accessible surface was determined using HOLLOW

(see Methods). **b**, The Slo2.2 constriction site is formed by Met333 from the inner helix of each of the four subunits. **c**, Electrostatic surface potential of the Slo2.2 ion conduction pathway, indicated by the colour scale.

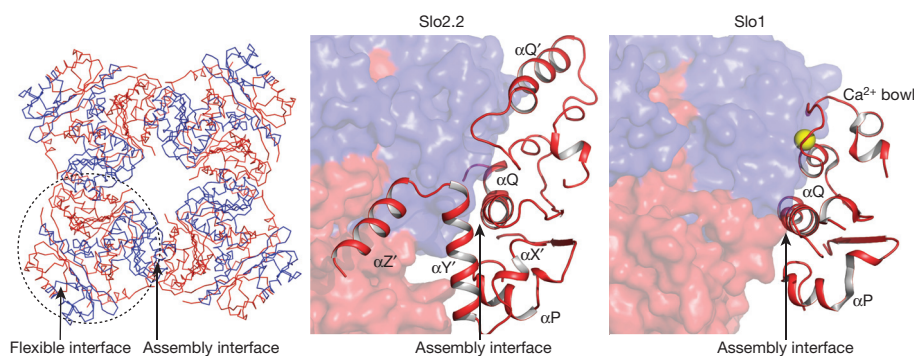
unique to Slo2 channels (Extended Data Fig. 1). Putative  $Na^+$ -binding-site amino acids, identified through mutagenesis, do not fall within the assembly interface<sup>42</sup>. If this means that  $Na^+$  in Slo2.2 activates from a different locus than does  $Ca^{2+}$  in Slo1, then we should not be too surprised—the prokaryotic  $K^+$  channel MthK contains a  $Ca^{2+}$ -activated gating ring in which  $Ca^{2+}$  binds near the flexible interface rather than the assembly interface<sup>43</sup>. Apparently the conformations of RCK-based gating rings can be modulated in a variety of ways.

A possible locus for  $Na^+$  binding was identified in two studies using mutagenesis<sup>42,44</sup>. Both studies found that Asp812 (or the corresponding residue in Slo2.1, Asp757) reduced  $Na^+$  activation. In one study, a second residue (His817 in Slo2.2) was also found to affect  $Na^+$  activation; however, position 817 is located 15 Å away from Asp812 and thus could not possibly form a single binding site along with Asp812 (ref. 42). Additional studies including a structure of Slo2.2 in the presence of  $Na^+$  are required to better understand  $Na^+$  activation in Slo2.2.

Crystal structures of isolated (from the TMD) Slo1 gating rings in the absence and presence of  $Ca^{2+}$  show that when  $Ca^{2+}$  binds to the  $Ca^{2+}$  bowl the N lobe of RCK1 undergoes a conformational change<sup>24,25</sup>. This change causes the four N lobes (one from each subunit) on the surface of the gating ring to expand like the petals of an opening flower (Extended Data Fig. 9). This expansion is illustrated by marking a residue (Lys343) on the N lobe that connects to S6 (the inner helix gate) through a linker (Extended Data Fig. 9). The distance between diagonally opposed Lys343 residues is approximately 81 Å and 94 Å in closed and opened Slo1 gating rings, respectively. The equivalent residue in Slo2.2 (Lys351) is 73 Å from its

diagonally opposed molecular symmetry mate, and the position of the N lobe is similar to that in the closed Slo1 gating ring. Thus, the Slo2.2 gating ring is closed, as is the inner helical gate to which it is connected. Cryo-EM density shows a weak connection between Lys351 on the N lobe and Lys337 on S6 (Fig. 6b). This connection supports the idea that expansion of the N lobes upon  $Na^+$  binding to the gating ring is mechanically connected to pore opening through the connecting linker. But the structure raises an additional mechanistic possibility. The interface that is formed between complementary-shaped surfaces on the TMD and the gating ring involves the RCK1 N lobes. If the N lobes expand when  $Na^+$  binds, as occurs when  $Ca^{2+}$  binds to the gating ring of Slo1, then the expansion would do more than pull on the linkers to S6; it would also presumably exert forces on the TMD through the protein–protein interface<sup>45</sup>. Such a mechanism of force transfer across the interface might explain why the S1–S4 domain in Slo2.2 is firmly attached through a large interface with the pore—so that conformational changes within the gating ring can be efficiently transmitted to the pore via the S1–S4 domain.

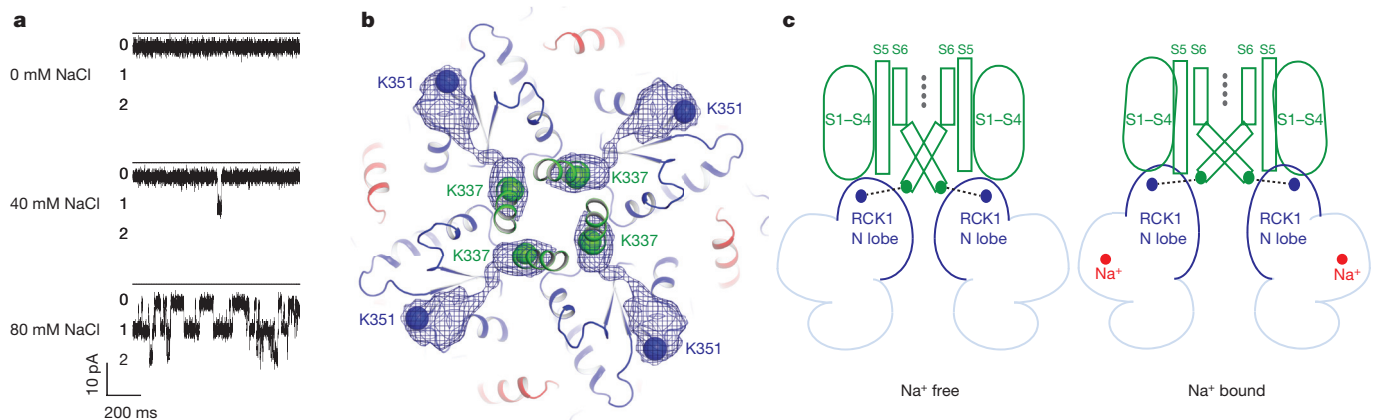
This study visualizes a  $Na^+$ -activated Slo2.2  $K^+$  channel in its ligand-free, closed conformation. The structure supports three important concepts concerning the function of this class of  $K^+$  channels: (1) a closed inner helix gate, analogous to that observed<sup>46</sup> or deduced<sup>47</sup> in other  $K^+$  channels, would prevent  $K^+$  conduction through restrictive van der Waal surfaces and an unfavourable electrostatic environment; (2) expansion of N lobes on the gating ring would open the pore through a polypeptide linker and through the protein–protein interface between the gating ring and the TMD; and (3) a massive electrostatic funnel facing the cytoplasm helps to explain why Slo channels are the highest-conductance  $K^+$  channels.



**Figure 5 | Slo2.2 gating ring.** Wire diagram of the Slo2.2 gating ring with the boundary of one subunit highlighted by the dashed circle (left). The flexible interface that mediates interactions between the RCK1 and RCK2 domains of each subunit and the assembly interface that mediates interactions between

adjacent subunits are shown. The assembly interfaces for Slo2.2 (middle) and Slo1 (right) with one subunit shown as a surface and the second shown as ribbons. The RCK1 domains are blue and the RCK2 domains are red. The  $Ca^{2+}$  ion resolved in the Slo1  $Ca^{2+}$  bowl is shown as a yellow sphere.





**Figure 6 | Slo2.2 gating.** **a**, Representative single channel recordings of Slo2.2 in planar lipid bilayers with symmetrical 0–80 mM NaCl and 150 mM KCl, when held at  $-80$  mV. **b**, Cryo-EM density connects Lys337 (the C-terminal residue of the TMD) to Lys351 (the N-terminal residue of the gating ring), the  $\alpha$  positions of which are highlighted by the green and blue spheres,

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 April; accepted 15 July 2015.

Published online 5 October 2015.

- Hille, B. *Ionic Channels of Excitable Membranes* 3rd edn 131–168 (Sinauer Associates, 2001).
- Bader, C. R., Bernheim, L. & Bertrand, D. Sodium-activated potassium current in cultured avian neurons. *Nature* **317**, 540–542 (1985).
- Dryer, S. E., Fujii, J. T. & Martin, A. R. A  $\text{Na}^+$ -activated  $\text{K}^+$  current in cultured brain stem neurons from chicks. *J. Physiol. (Lond.)* **410**, 283–296 (1989).
- Haimann, C. & Bader, C. R. Sodium-activated potassium channel in avian sensory neurons. *Cell Biol. Int. Rep.* **13**, 1133–1139 (1989).
- Kameyama, M. *et al.* Intracellular  $\text{Na}^+$  activates a  $\text{K}^+$  channel in mammalian cardiac cells. *Nature* **309**, 354–356 (1984).
- Schwindt, P. C., Spain, W. J. & Crill, W. E. Long-lasting reduction of excitability by a sodium-dependent potassium current in cat neocortical neurons. *J. Neurophysiol.* **61**, 233–244 (1989).
- Yan, Y., Yang, Y., Bian, S. & Sigworth, F. J. Expression, purification and functional reconstitution of slack sodium-activated potassium channels. *J. Membr. Biol.* **245**, 667–674 (2012).
- Yuan, A. *et al.* The sodium-activated potassium channel is encoded by a member of the *Slo* gene family. *Neuron* **37**, 765–773 (2003).
- Bhattacharjee, A., Gan, L. & Kaczmarek, L. K. Localization of the Slack potassium channel in the rat central nervous system. *J. Comp. Neurol.* **454**, 241–254 (2002).
- Kaczmarek, L. K. *et al.* Regulation of the timing of MNTB neurons by short-term and long-term modulation of potassium channels. *Hear. Res.* **206**, 133–145 (2005).
- Wallén, P. *et al.* Sodium-dependent potassium channels of a Slack-like subtype contribute to the slow afterhyperpolarization in lamprey spinal neurons. *J. Physiol. (Lond.)* **585**, 75–90 (2007).
- Yang, B., Desai, R. & Kaczmarek, L. K. Slack and Slick  $\text{K}_{\text{Na}}$  channels regulate the accuracy of timing of auditory neurons. *J. Neurosci.* **27**, 2617–2627 (2007).
- Allen, A. S. *et al.* De novo mutations in epileptic encephalopathies. *Nature* **501**, 217–221 (2013).
- Barcia, G. *et al.* De novo gain-of-function *KCN1* channel mutations cause malignant migrating partial seizures of infancy. *Nature Genet.* **44**, 1255–1259 (2012).
- Ishii, A. *et al.* A recurrent *KCN1* mutation in two sporadic cases with malignant migrating partial seizures in infancy. *Gene* **531**, 467–471 (2013).
- McTague, A. *et al.* Migrating partial seizures of infancy: expansion of the electroclinical, radiological and pathological disease spectrum. *Brain* **136**, 1578–1591 (2013).
- Vanderver, A. *et al.* Identification of a novel de novo p.Phe932Ile *KCN1* mutation in a patient with leukoencephalopathy and severe epilepsy. *Pediatr. Neurol.* **50**, 112–114 (2014).
- Heron, S. E. *et al.* Missense mutations in the sodium-gated potassium channel gene *KCN1* cause severe autosomal dominant nocturnal frontal lobe epilepsy. *Nature Genet.* **44**, 1188–1190 (2012).
- Martin, H. C. *et al.* Clinical whole-genome sequencing in severe early-onset epilepsy reveals new genes and improves molecular diagnosis. *Hum. Mol. Genet.* **23**, 3200–3211 (2014).
- Huang, F. *et al.* TMEM16C facilitates  $\text{Na}^+$ -activated  $\text{K}^+$  currents in rat sensory neurons and regulates pain processing. *Nature Neurosci.* **16**, 1284–1290 (2013).
- Lu, R. *et al.* Slack channels expressed in sensory neurons control neuropathic pain in mice. *J. Neurosci.* **35**, 1125–1135 (2015).

respectively. The density is filtered to 5 Å. **c**, Model for Slo2.2 channel activation.  $\text{Na}^+$  binding to the gating ring results in an outward movement of the RCK1 N lobes, which pull apart the inner S6 helices and push up the cytoplasmic ends of S4 and S5, thereby opening the pore.

- Lu, S., Das, P., Fadool, D. A. & Kaczmarek, L. K. The Slack sodium-activated potassium channel provides a major outward current in olfactory neurons of *Kv1.3<sup>-/-</sup>* super-smeller mice. *J. Neurophysiol.* **103**, 3311–3319 (2010).
- Paulais, M., Lachheb, S. & Teulon, J. A  $\text{Na}^+$ - and  $\text{Cl}^-$ -activated  $\text{K}^+$  channel in the thick ascending limb of mouse kidney. *J. Gen. Physiol.* **127**, 205–215 (2006).
- Wu, Y., Yang, Y., Ye, S. & Jiang, Y. Structure of the gating ring from the human large-conductance  $\text{Ca}^{2+}$ -gated  $\text{K}^+$  channel. *Nature* **466**, 393–397 (2010).
- Yuan, P., Leonetti, M. D., Hsiung, Y. & MacKinnon, R. Open structure of the  $\text{Ca}^{2+}$  gating ring in the high-conductance  $\text{Ca}^{2+}$ -activated  $\text{K}^+$  channel. *Nature* **481**, 94–97 (2012).
- Yuan, P., Leonetti, M. D., Pico, A. R., Hsiung, Y. & MacKinnon, R. Structure of the human BK channel  $\text{Ca}^{2+}$ -activation apparatus at 3.0 Å resolution. *Science* **329**, 182–186 (2010).
- Garg, P., Gardner, A., Garg, V. & Sanguinetti, M. C. Structural basis of ion permeation gating in Slo2.1  $\text{K}^+$  channels. *J. Gen. Physiol.* **142**, 523–542 (2013).
- Wilkens, C. M. & Aldrich, R. W. State-independent block of BK channels by an intracellular quaternary ammonium. *J. Gen. Physiol.* **128**, 347–364 (2006).
- Zhou, Y., Xia, X. M. & Lingle, C. J. Cysteine scanning and modification reveal major differences between BK channels and Kv channels in the inner pore region. *Proc. Natl Acad. Sci. USA* **108**, 12161–12166 (2011).
- Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent  $\text{K}^+$  channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a  $\text{K}^+$  channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- Joiner, W. J. *et al.* Formation of intermediate-conductance calcium-activated potassium channels by interaction of Slack and Slo subunits. *Nature Neurosci.* **1**, 462–469 (1998).
- Zhang, X., Zeng, X. & Lingle, C. J. Slo3  $\text{K}^+$  channels: voltage and pH dependence of macroscopic currents. *J. Gen. Physiol.* **128**, 317–336 (2006).
- Dai, L., Garg, V. & Sanguinetti, M. C. Activation of Slo2.1 channels by niflumic acid. *J. Gen. Physiol.* **135**, 275–295 (2010).
- Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family  $\text{K}^+$  channel. *Science* **309**, 897–903 (2005).
- Tao, X., Avalos, J. L., Chen, J. & MacKinnon, R. Crystal structure of the eukaryotic strong inward-rectifier  $\text{K}^+$  channel Kir2.2 at 3.1 Å resolution. *Science* **326**, 1668–1674 (2009).
- Whorton, M. R. & MacKinnon, R. Crystal structure of the mammalian GIRK2  $\text{K}^+$  channel and gating regulation by G proteins,  $\text{PIP}_2$ , and sodium. *Cell* **147**, 199–208 (2011).
- Guo, D. & Lu, Z. Interaction mechanisms between polyamines and IRK1 inward rectifier  $\text{K}^+$  channels. *J. Gen. Physiol.* **122**, 485–500 (2003).
- Heginbotham, L. & MacKinnon, R. Conduction properties of the cloned Shaker  $\text{K}^+$  channel. *Biophys. J.* **65**, 2089–2096 (1993).
- Budelli, G., Geng, Y., Butler, A., Magleby, K. L. & Salkoff, L. Properties of Slo1  $\text{K}^+$  channels with and without the gating ring. *Proc. Natl Acad. Sci. USA* **110**, 16657–16662 (2013).
- Leonetti, M. D., Yuan, P., Hsiung, Y. & MacKinnon, R. Functional and structural analysis of the human Slo3 pH- and voltage-gated  $\text{K}^+$  channel. *Proc. Natl Acad. Sci. USA* **109**, 19274–19279 (2012).
- Zhang, Z., Rosenhouse-Dantsker, A., Tang, Q. Y., Noskov, S. & Logothetis, D. E. The RCK2 domain uses a coordination site present in Kir channels to confer sodium sensitivity to Slo2.2 channels. *J. Neurosci.* **30**, 7554–7562 (2010).
- Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).

44. Thomson, S. J., Hansen, A. & Sanguinetti, M. C. Identification of the intracellular Na<sup>+</sup> sensor in Slo2.1 potassium channels. *J. Biol. Chem.* **290**, 14528–14535 (2015).
45. Yang, H. *et al.* Activation of Slo1 BK channels by Mg<sup>2+</sup> coordinated between the voltage sensor and RCK1 domains. *Nature Struct. Mol. Biol.* **15**, 1152–1159 (2008).
46. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **280**, 69–77 (1998).
47. Webster, S. M., Del Camino, D., Dekker, J. P. & Yellen, G. Intracellular gate opening in Shaker K<sup>+</sup> channels defined by high-affinity metal bridges. *Nature* **428**, 864–868 (2004).

**Acknowledgements** We thank Z. Yu and J. de la Cruz at the Howard Hughes Medical Institute Janelia Cryo-EM facility for assistance in data collection, S. Harrison and S. Jenni for assistance with Phenix refinement of cryo-EM density maps, and members of the MacKinnon laboratory for discussions. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. This work was supported in part by GM43949. R.K.H. is a Howard Hughes Medical Institute postdoctoral fellow of the Helen Hay Whitney Foundation and T.W. and R.M. are investigators of the Howard Hughes Medical Institute.

**Author Contributions** R.K.H. performed the experiments. P.Y. provided assistance with protein expression and purification. Z.L. aided with sample preparation and data collection. Y.H. provided assistance with protein expression. T.W. aided with initial model generation and map interpretation. R.K.H. and R.M. designed the experiments and analysed the results. R.K.H. and R.M. prepared the manuscript with input from all co-authors.

**Author Information** The 3D cryo-EM density maps of Slo2.2 with low-pass filter and amplitude modification have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-3062 (Slo2.2 whole channel), EMD-3063 (Slo2.2 gating ring) and EMD-3064 (Slo2.2 TMD). Atomic coordinates for the atomic model of full-length Slo2.2, Slo2.2 gating ring and Slo2.2 TMD have been deposited in the Protein Data Bank under accession numbers 5A6E, 5A6F and 5A6G, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M. ([mackinn@rockefeller.edu](mailto:mackinn@rockefeller.edu)).

## METHODS

No statistical methods were used to predetermine sample size.

**Expression and purification.** A synthetic gene fragment encoding residues 1 to 1201 of chicken Slo2.2 (GI:73920088) was purchased from Bio Basic Inc. The resulting fragment was cloned into a modified pFastbac vector (Invitrogen) containing green fluorescent protein and a 1D4 antibody recognition sequence (TETSQVAPA) on the C terminus. Sf9 (*Spodoptera frugiperda*) cells infected with the baculovirus were cultured at 27 °C for 72 h in supplemented Grace's insect cell medium (Invitrogen). Cells were washed with ice-cold phosphate-buffered saline and extracted for 3 h at 4 °C with buffer containing 50 mM Hepes pH 7.4, 300 mM KCl and 40 mM dodecyl- $\beta$ -D-maltopyranoside (DDM) in the presence of a protease inhibitor cocktail (2  $\mu$ g ml<sup>-1</sup> leupeptin, 2  $\mu$ g ml<sup>-1</sup> aprotinin, 2  $\mu$ g ml<sup>-1</sup> pepstatin A, 1 mM benzamidine, 100  $\mu$ g ml<sup>-1</sup> 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride and 100  $\mu$ M phenylmethane sulphonylfluoride). The insoluble fraction was removed by centrifugation at 35,000g for 45 min at 4 °C and the remaining soluble fraction was incubated with 1D4-affinity resin pre-equilibrated with 20 mM Hepes pH 7.4, 300 mM KCl and 4 mM DDM. The suspension was mixed for 5 h at 4 °C. Beads were collected on a column by gravity and then washed with 10 column volumes of wash buffer (20 mM Hepes pH 7.4, 300 mM KCl, 4 mM DDM and 0.1 mg ml<sup>-1</sup> 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE):1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol (POPG) (3:1 w/w)). The protein was digested with PreScission protease (20:1 w/w ratio) on the column overnight at 4 °C to remove the affinity tag and then eluted with two column volumes of wash buffer. Concentrated protein was further purified by size-exclusion chromatography on a Superose 6 (GE Life Sciences) column in 20 mM Hepes pH 7.4, 300 mM KCl, 1.5 mM DDM and 0.05 mg ml<sup>-1</sup> POPE:POPG (3:1 w/w). Peak fractions were pooled and concentrated to 3–5 mg ml<sup>-1</sup> for cryo-EM analysis.

**Electron microscopy sample preparation and imaging.** 3.5  $\mu$ l of purified channel was pipetted onto glow-discharged copper Quantifoil R 1.2/1.3 holey carbon grids (Quantifoil). Grids were blotted for 3.5 s at about 85% humidity and flash frozen in liquid-nitrogen-cooled liquid ethane using an FEI Vitrobot Mark IV (FEI). Grids were transferred to an FEI Titan Krios electron microscope operating at an acceleration voltage of 300 keV. Images were recorded in an automated fashion on a Gatan K2 Summit (Gatan) detector set to super-resolution counting mode with a super-resolution pixel size of 0.52 Å using SerialEM<sup>48</sup>. Inelastically scattered electrons were filtered by a Gatan Imaging Filter with a slit width of 30 eV (Gatan). Images were recorded for 5 s with a subframe exposure time of 200 ms and a dose of approximately eight electrons per physical pixel (1.04 Å at the image plane) for a total accumulated dose of approximately 40 electrons per square ångström on the specimen over 25 subframes (approximately 1.5 electrons per square ångström per subframe).

**Image processing and map calculation.** Dose-fractionated super-resolution images were 2 × 2 binned (resulting in a pixel size of 1.04 Å) for whole-frame motion correction with motioncorr<sup>49</sup>. Following motion correction, the sum of the 25 aligned subframes in the image stack was 4 × 4 binned (pixel size of 4.16 Å) for manual particle selection. 40,643 particles were manually selected in Boxer<sup>50</sup> from 2,243 images and then extracted from the motion-corrected images in RELION<sup>51</sup>. The effects of the contrast transfer function were estimated by ctffind3<sup>52</sup> and CTER<sup>53</sup>. 22,154 particles were subjected to iterative stable alignment and classification<sup>54</sup>, specifying 200 particles per group and a pixel error threshold of 2; 6 generations yielded a set of 642 stable class averages. The classes were inspected manually and those whose images contained portions of more than one particle or that had other obvious defects were discarded. The remaining 472 classes were used to generate an initial model with the multi\_shc command in SPARX<sup>55</sup> with and without imposing C<sub>4</sub> symmetry. The results were similar for both cases, so the symmetrized version was used as the initial model for 3D refinement (Extended Data Fig. 3c.).

The initial set of 40,643 particles was subjected to 2D classification using RELION, from which particles classified into low-abundance classes were removed, resulting in a stack of 35,586 particles. These 35,586 particles were segregated into six classes by RELION's 3D classification algorithm using the initial model generated in SPARX. Four of the six classes appeared similar by visual inspection in UCSF Chimera<sup>56</sup> and were combined into a stack of 24,231 particles for 3D refinement in RELION. Particle polishing was applied to the particles using RELION 1.3 to correct for individual particle movements and for attenuation of the high-resolution terms due to rapid movement of the initial few frames and radiation damage in the later frames<sup>57</sup>, yielding a map that achieved a resolution of 4.5 Å as assessed by Fourier shell correlation (FSC) of independently refined half data sets and using the 0.143 cut-off criterion<sup>58</sup>. Local resolution estimates calculated using ResMap demonstrated that the periphery of the TMD was poorly ordered compared to the pore and the gating ring<sup>59</sup>.

3D maximum-likelihood classification of the 24,231 particles was performed using FREALIGN<sup>60</sup>, starting from the orientation parameters determined by RELION. Five subclasses were specified and the resolution was limited to 8.8 Å for classification and 6.6 Å for refinement to minimize the effects of high-resolution overfitting. Maps of each of the subclasses were calculated at 7 Å and were superimposed on the basis of their gating rings using UCSF Chimera. The orientations of the TMDs varied by up to 7°.

To minimize the effects of the rotational variability, soft masks were generated (five-pixel cushion and five-pixel cosine edge fall-off) for both the gating ring and the TMD and applied to the reconstructions during refinement<sup>61</sup>. The gating-ring mask was applied at the end of every round of refinement in RELION, resulting in a map that reached a resolution of 4.2 Å. The variation of the TMD made it difficult to align the entire data set, so instead the two most similar classes from the 3D classification were combined, yielding a subset of 11,303 images. The angular parameters of these images were determined by FREALIGN using an iterative process in which two refinement cycles were run with the mask followed by two without the mask. This iterative procedure of refining with and without mask application was necessary, owing to the small mass of the TMD relative to that of the whole channel. The map of the transmembrane region achieved a final resolution of 5.2 Å by FSC.

**Model building.** The structure of the closed Slo1 gating ring<sup>24</sup> (Protein Data Bank (PDB) code 3NAF) was docked into the gating-ring map using UCSF Chimera and then manually rebuilt in Coot<sup>62</sup> to fit the density. For the regions lacking homology with Slo1, secondary structure elements were first built into the density map and then connected with manually built loops. Once the backbone was traced, the sequence was registered using a combination of identification of large side chains and sequence homology with Slo1. The structure of the Kv chimaera  $\alpha$ -subunit<sup>30</sup> (PDB code 2R9R) was docked into the TMD map using UCSF Chimera using the selectivity filter and pore helix as a reference. The positions and tilts of S5 and S6 were adjusted by moving the helices as rigid domains to fit the density, and then the loops connecting them were manually rebuilt. The sequence was registered using the presence of large side chains, the proline kink and sequence homology among K<sup>+</sup> channels. The S1–S4 domain was built as six polyalanine helices in Coot (S0, 15 residues; S1, 28 residues; S1–S2, 17 residues; S2 and S3, 53 residues; S4, 26 residues). The helices were assigned on the basis of homology with Kv chimaera, and roughly agree in length with those predicted by hydropathy analysis. A model of the full channel was generated by manually docking the independently determined structures for the gating ring and the TMD into the map of the full channel.

**Reciprocal space refinement.** The masked cryo-EM density maps were translated to a new unit cell that extended 10 Å from the model in all directions, and then solvent flattened by creating a mask that extends 3 Å away from the model and setting the density outside of this mask to approximate the solvent content of a protein crystal. Reciprocal-space structure factors were calculated from the translated and solvent-flattened maps using FFT<sup>63</sup>. The phases were blurred in a resolution-dependent fashion by setting the figure of merit equal to the FSC and subsequently used to calculate Hendrickson–Lattman coefficients. Coordinate and B-factor refinement of the gating-ring and TMD models were performed in Phenix<sup>64</sup>, using X-ray scattering factors against a maximum likelihood with an experimental phase probability distribution target function, with 5% of the reflections excluded as a free set to monitor overfitting. Secondary structure, geometric, B-factor and Ramachandran restraints were used throughout refinement. Rigid-body refinement of the full channel model was performed in Phenix<sup>64</sup> using X-ray scattering factors against a maximum likelihood with an experimental phase probability distribution target function, with 5% of the reflections excluded as a free set to monitor overfitting.

Overfitting of the models during refinement was also evaluated using a method previously described<sup>61</sup>. Briefly, all atoms of the refined models were randomly displaced by 0.1 Å and then subjected to one round of coordinate and B-factor refinement against phases and amplitudes calculated from one of the independently determined half maps. Following this refinement cycle, the FSC was calculated between the resultant model and both of the half maps, as well as the full map. Structure figures were prepared with UCSF Chimera<sup>56</sup>, Pymol (Pymol version 1.7.2 Schrödinger, LLC), HOLLOW<sup>65</sup> and APBS<sup>66,67</sup>. All structure calculations were performed using software compiled by SGrid<sup>68</sup>.

**Reconstitution and electrophysiological recordings from planar lipid bilayers.** Purified channels were reconstituted into octyl maltoside (Anatrace)-solubilized 3:1 w/w POPE:POPG lipid vesicles as described<sup>69</sup>. Detergent was removed by dialysis for 5 d against a detergent-free buffer containing 10 mM HEPES-KOH pH 7.5, 450 mM KCl and 2 mM dithiothreitol at 4 °C, with daily buffer exchanges. After 5 d, all residual detergent was removed by incubating the reconstituted channels with Bio-Beads (Bio-Rad) for 2 h at room temperature. The reconsti-



tuted channels were aliquoted and flash frozen into liquid nitrogen before storage at  $-80^{\circ}\text{C}$ .

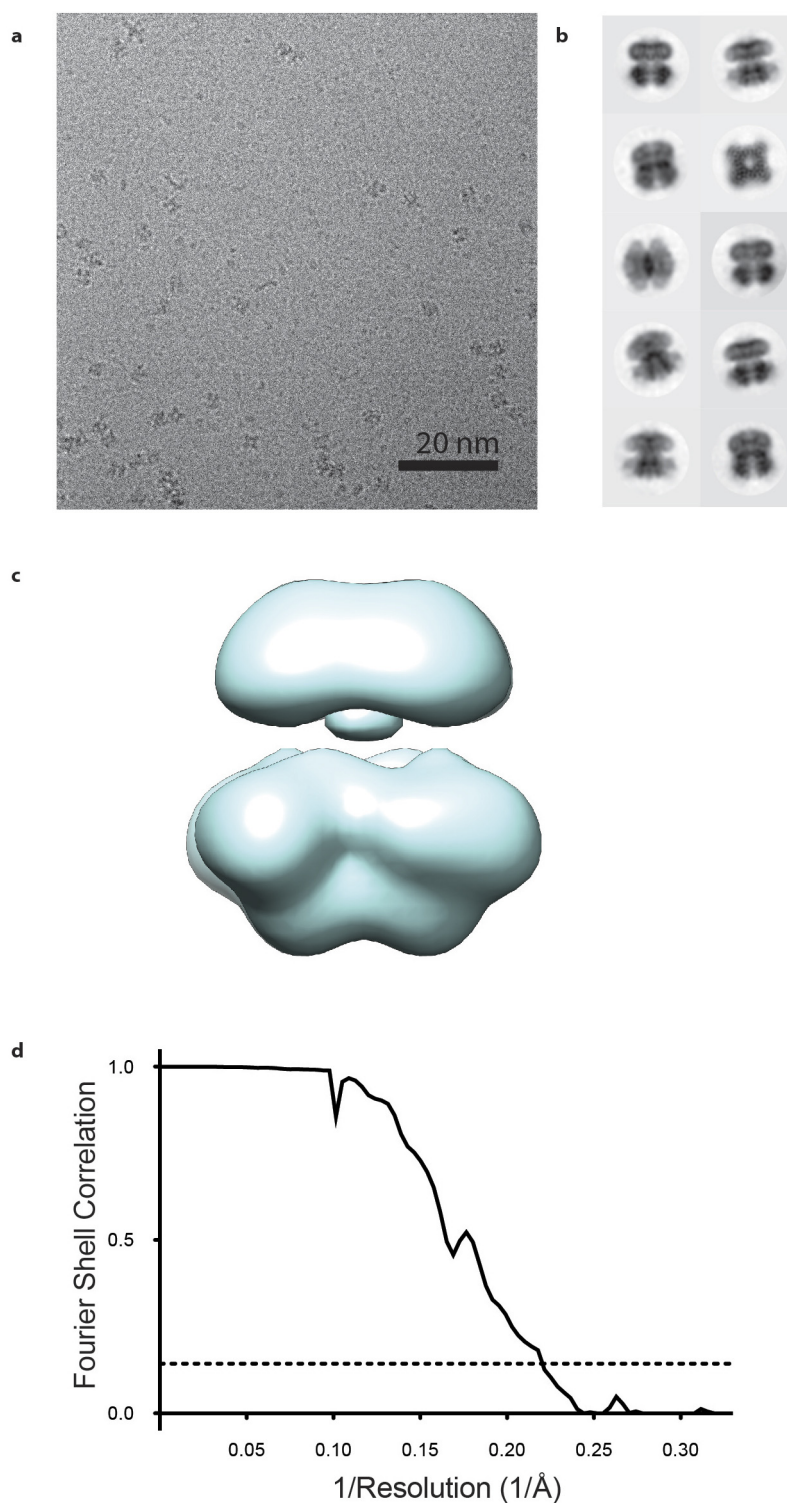
Planar lipid bilayer experiments were performed as described previously<sup>70,71</sup>. Lipids of desired compositions were prepared by dissolving argon-dried lipids in decane to a final concentration of  $20\text{ mg ml}^{-1}$ . Lipid solutions were painted over a  $300\text{-}\mu\text{m}$  hole in a polystyrene partition that separated the two chambers to form the planar lipid bilayer. Reconstituted channels were pipetted onto the chamber side of the bilayer after thinning of a planar lipid bilayer had been detected via monitoring of electrical capacitance. All recordings were performed using the voltage-clamp method in whole-cell mode. Analogue signals were filtered at  $1\text{ kHz}$  using a low-pass Bessel filter on an Axopatch 200B amplifier (Molecular Devices) in whole-cell mode and digitized at  $10\text{ kHz}$  using a Digidata 1400A analogue-to-digital converter (Molecular Devices). The pClamp software suite (Molecular Devices) was used to control membrane voltage and record current. All electrophysiological recordings were replicated in at least three independent experiments.

48. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
49. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
50. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
51. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
52. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
53. Penczek, P. A. *et al.* CTER-rapid estimation of CTF parameters with error assessment. *Ultramicroscopy* **140**, 9–19 (2014).
54. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237–247 (2012).
55. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
56. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
57. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
58. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
59. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
60. Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388 (2013).
61. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
62. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
63. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
64. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
65. Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).
66. Dolinsky, T. J. *et al.* PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).
67. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
68. Morin, A. *et al.* Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
69. Schmidt, D., Jiang, Q. X. & MacKinnon, R. Phospholipids and the origin of cationic gating charges in voltage sensors. *Nature* **444**, 775–779 (2006).
70. Miller, C. (ed.) *Ion Channel Reconstitution* (Plenum, 1986).
71. Ruta, V., Jiang, Y., Lee, A., Chen, J. & MacKinnon, R. Functional analysis of an archaeobacterial voltage-dependent  $\text{K}^+$  channel. *Nature* **422**, 180–185 (2003).
72. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).



**Extended Data Figure 1 | Sequence alignment of Slo channels.** **a**, Sequence alignment of chicken Slo2.2 with human Slo2.2 and human Slo2.1. **b**, Predicted position of transmembrane helices in Slo2.2 S1–S4 domain on the basis of hydropathy analysis using Jpred 4 (ref. 72). **c, d**, Structure-based sequence

alignment of chicken Slo2.2 TMD with rat Kv chimera (c) and chicken Slo2.2 gating ring with human Slo1 gating ring (d). Helices are blue and  $\beta$ -strands are red.

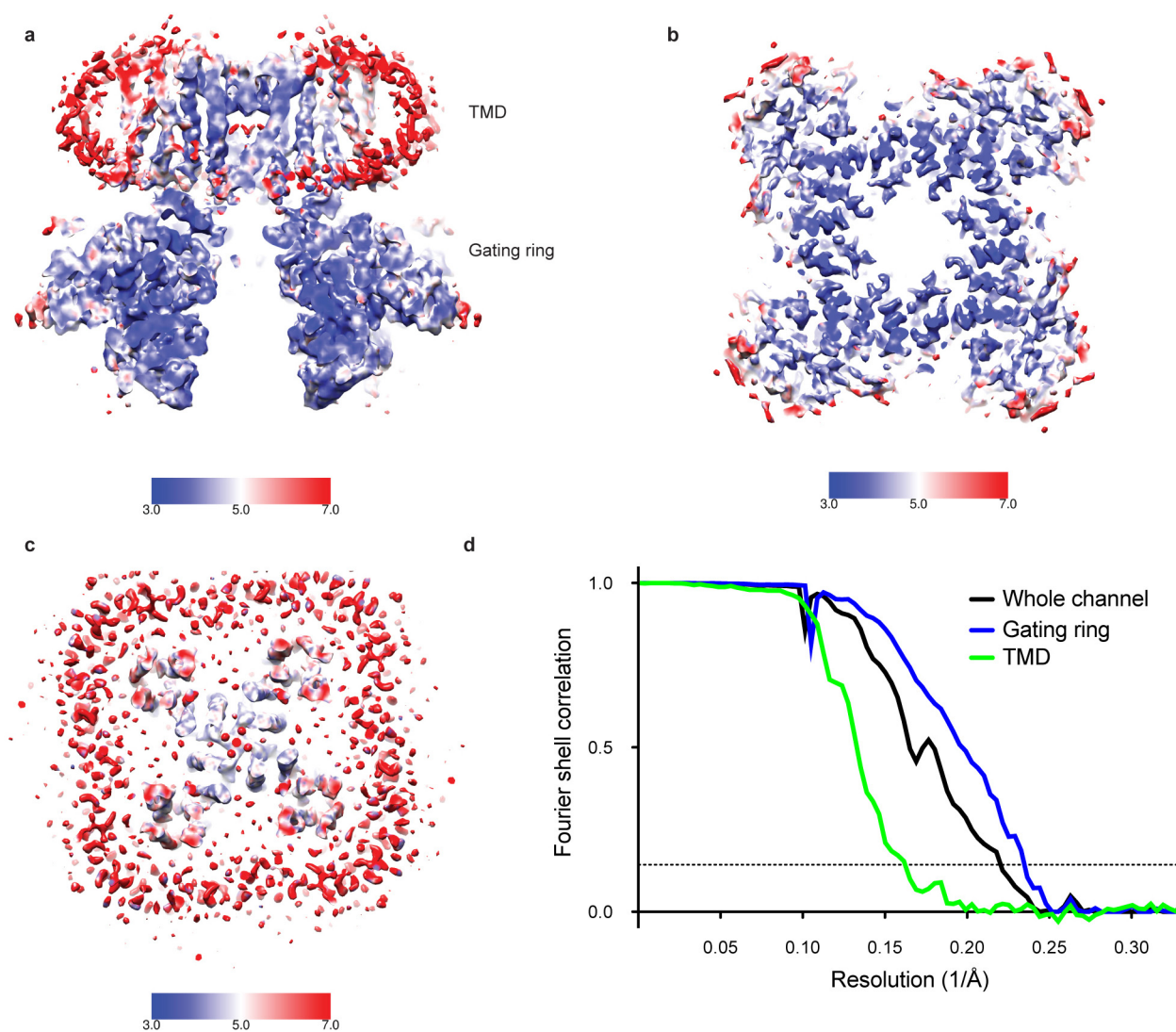


# Extended Data Figure 2 | Full channel 3D reconstruction of chicken Slo2.2.

**a**, Representative micrograph of detergent- and lipid-solubilized Slo2.2 in vitreous ice. **b**, Selected 2D class averages. **c**, *Ab initio* model of Slo2.2. **d**, FSC

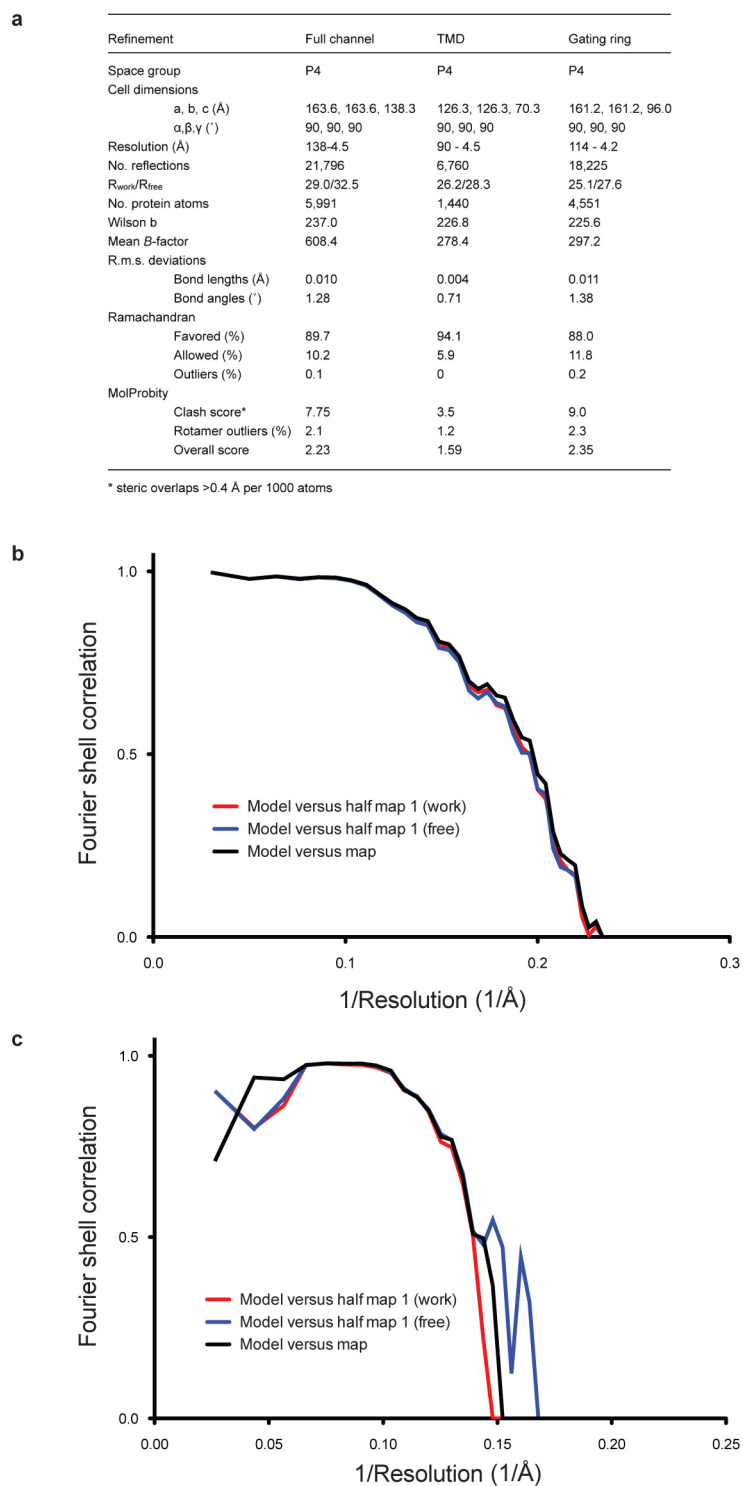
curve of the full channel reconstruction with the nominal resolution estimated to be 4.5 Å on the basis of the FSC = 0.143 (dashed line) cut-off criterion.





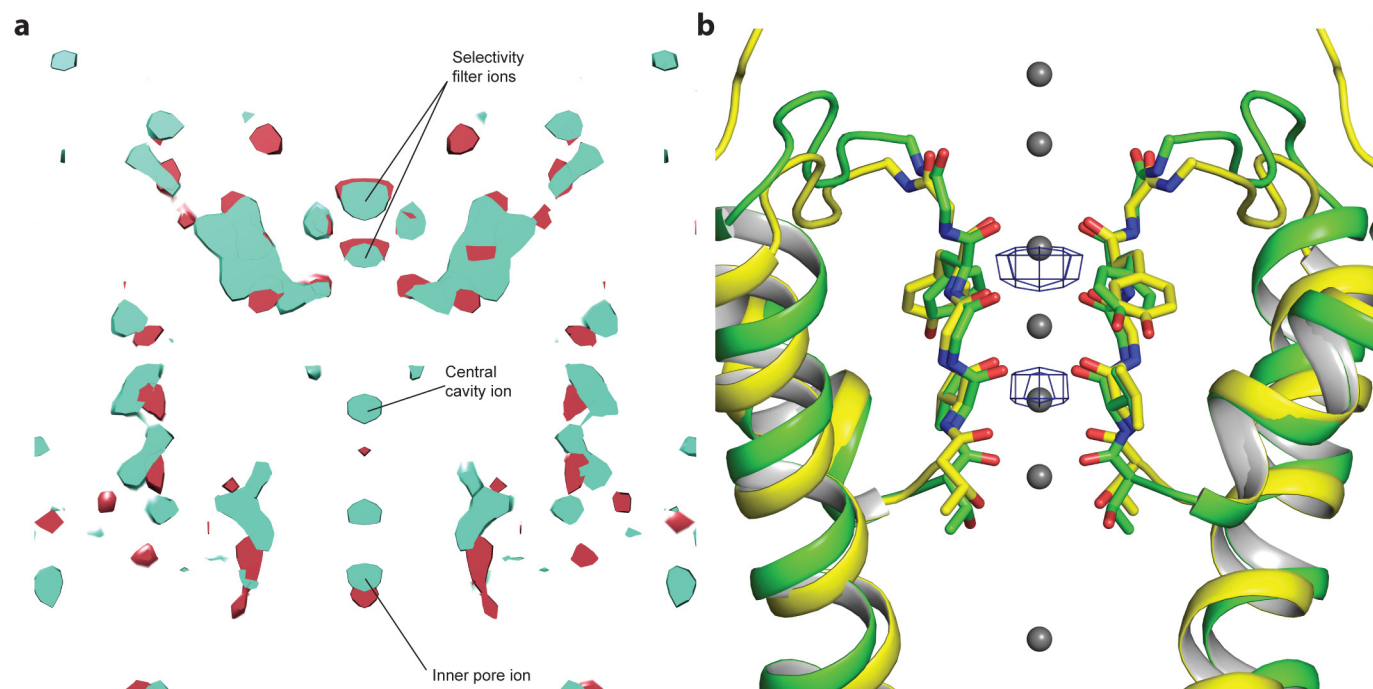
**Extended Data Figure 3 | Focused refinement of the gating ring and the TMD.** **a**, 3D density map of the full channel reconstruction, coloured according to local resolution (in Å). **b**, **c**, 3D density map calculated following focused refinement using a mask to only include the gating ring (**b**) and the

TMD (**c**), coloured according to local resolution (in Å). **d**, FSC of the full channel reconstruction (estimated resolution of 4.5 Å), the gating-ring-focused refinement reconstruction (4.2 Å) and the TMD-focused refinement reconstruction (5.2 Å).



**Extended Data Figure 4 | Validation of the Slo2.2 model.** **a**, Refinement statistics for the Slo2.2 full channel, TMD and gating-ring models. **b**, **c**, FSC curves for cross-validation of the refined gating ring (**b**) and TMD (**c**) models. The black curves are the refined model compared to the full data set, the

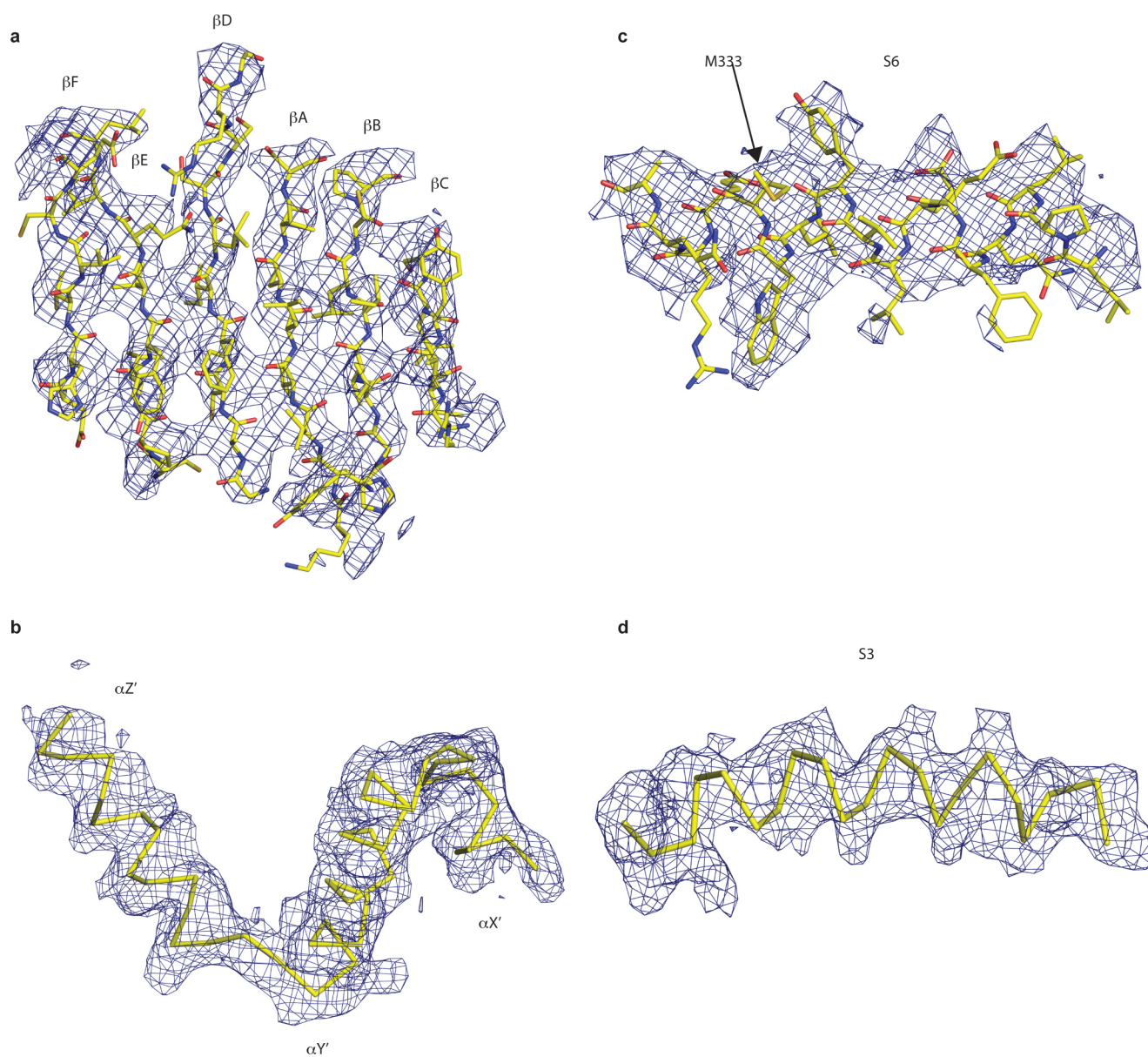
red curves are the refined model compared to half map 1 (used during test refinement) and the blue curves are the refined model compared to half map 2 (not used during test refinement).



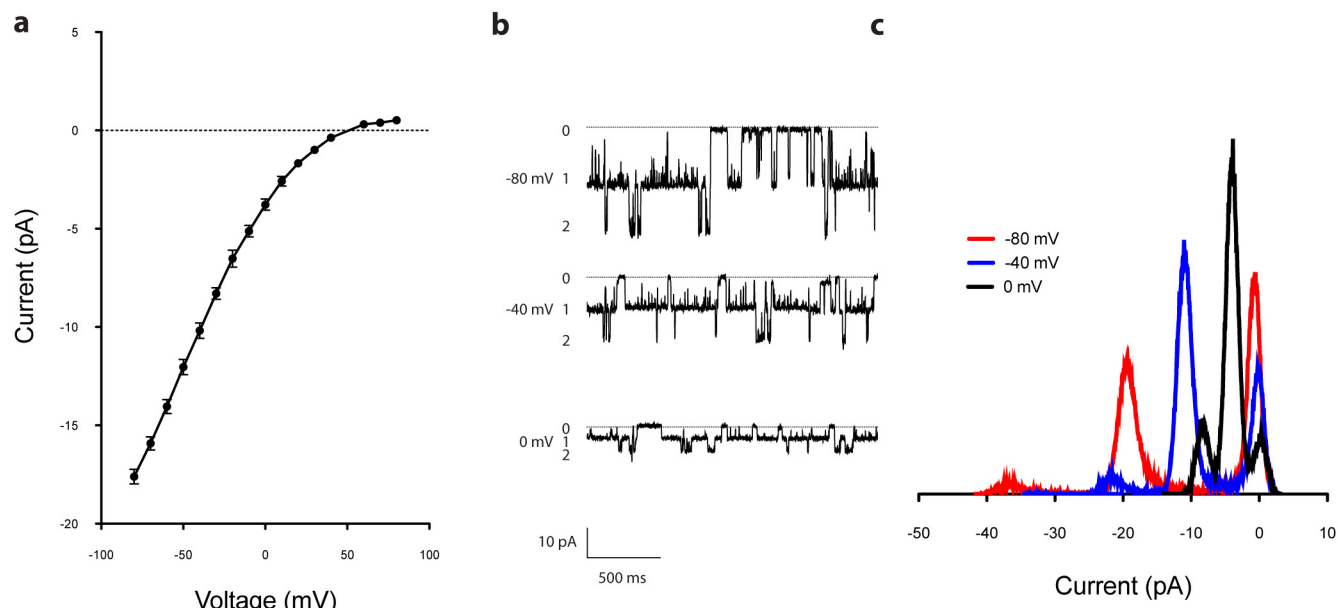
**Extended Data Figure 5 |  $K^+$  ions in Slo2.2.** **a**, Central section of the density maps of the two independently calculated half maps (coloured in green and red) with densities corresponding to  $K^+$  ions labelled. **b**, Superposition of

the Slo2.2 selectivity filter (green) with KcsA (PDB code 1K4C) selectivity filter (yellow). Density peaks resolved in the Slo2.2 selectivity filter at  $6.5\sigma$  are shown as blue meshes.  $K^+$  ions resolved in KcsA are shown as grey spheres.



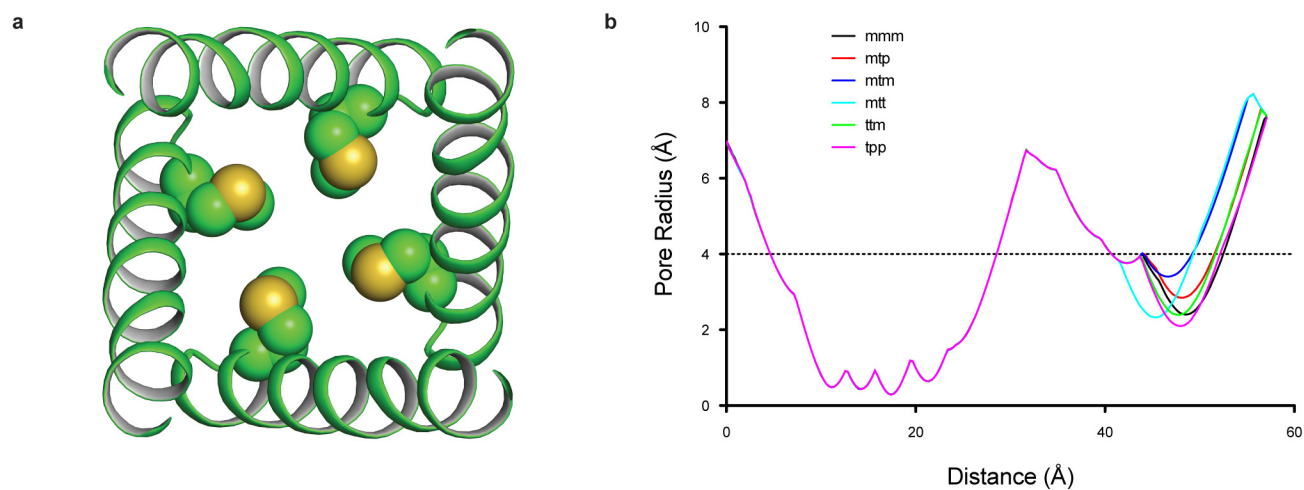


**Extended Data Figure 6 | Representative segments of the cryo-EM density map.** a–d, Selected regions of the gating-ring density (a, b) and the TMD density (c, d) maps with the refined model.



**Extended Data Figure 7 | Single channel conductance of Slo2.2.** **a**, Single channel current–voltage relationship (mean  $\pm$  s.e.m.) for Slo2.2 in planar lipid bilayers. Single channel conductance is about 200 pS. **b**, Representative recordings of Slo2.2 held at  $-80$  mV,  $-40$  mV and  $0$  mV in planar lipid

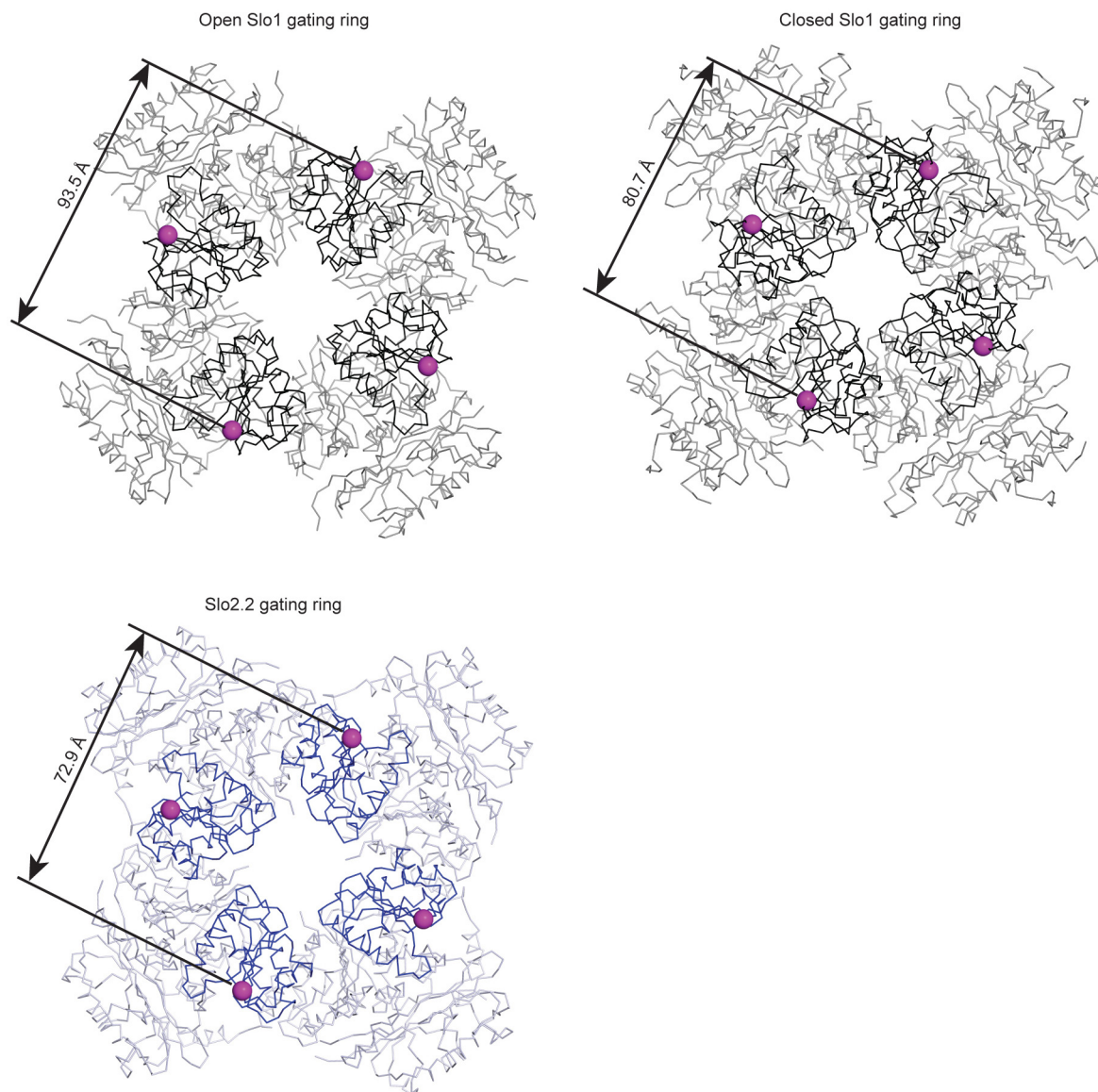
bilayers. Chamber solution contained 135 mM NaCl and 15 mM KCl, and cup solution contained 150 mM KCl. **c**, Histogram of Slo2.2 currents when held at  $-80$  mV,  $-40$  mV and  $0$  mV, as labelled.



**Extended Data Figure 8 | Inner helix gate.** **a**, Ribbon diagram of the Slo2.2 pore with Met333 side chains modelled as spheres. **b**, Pore radius plot as a function of distance from the extracellular surface for Slo2.2 with Met333

modelled as each of the six most frequently observed rotamers, as labelled. For distances less than about 40 Å, the curves coincide.





**Extended Data Figure 9 | Slo2.2 gating ring is in a closed conformation.** Wire diagrams of Slo1 gating ring in the open (top left) and closed (top right) conformations. The mobile RCK1 N lobe is black and the rest of the gating ring is grey. The N-terminal residue of the gating ring, Lys343, is shown as a

pink sphere. Wire diagram of the Slo2.2 gating ring (bottom) with the RCK1 N-lobe blue and the rest of the gating ring light blue. The N-terminal residue of the gating ring, Lys351, is shown as a pink sphere.

Extended Data Table 1 | 3D reconstructions of chicken Slo2.2 by cryo-EM

Refinement	Full channel	TMD	Gating ring
Space group	P4	P4	P4
Cell dimensions			
a, b, c (Å)	163.6, 163.6, 138.3	126.3, 126.3, 70.3	161.2, 161.2, 96.0
$\alpha, \beta, \gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	138-4.5	90 - 4.5	114 - 4.2
No. reflections	21,796	6,760	18,225
$R_{\text{work}}/R_{\text{free}}$	29.0/32.5	26.2/28.3	25.1/27.6
No. protein atoms	5,991	1,440	4,551
Wilson b	237.0	226.8	225.6
Mean <i>B</i> -factor	608.4	278.4	297.2
R.m.s. deviations			
Bond lengths (Å)	0.010	0.004	0.011
Bond angles (°)	1.28	0.71	1.38
Ramachandran			
Favored (%)	89.7	94.1	88.0
Allowed (%)	10.2	5.9	11.8
Outliers (%)	0.1	0	0.2
MolProbity			
Clash score*	7.75	3.5	9.0
Rotamer outliers (%)	2.1	1.2	2.3
Overall score	2.23	1.59	2.35

\* steric overlaps &gt;0.4 Å per 1000 atoms

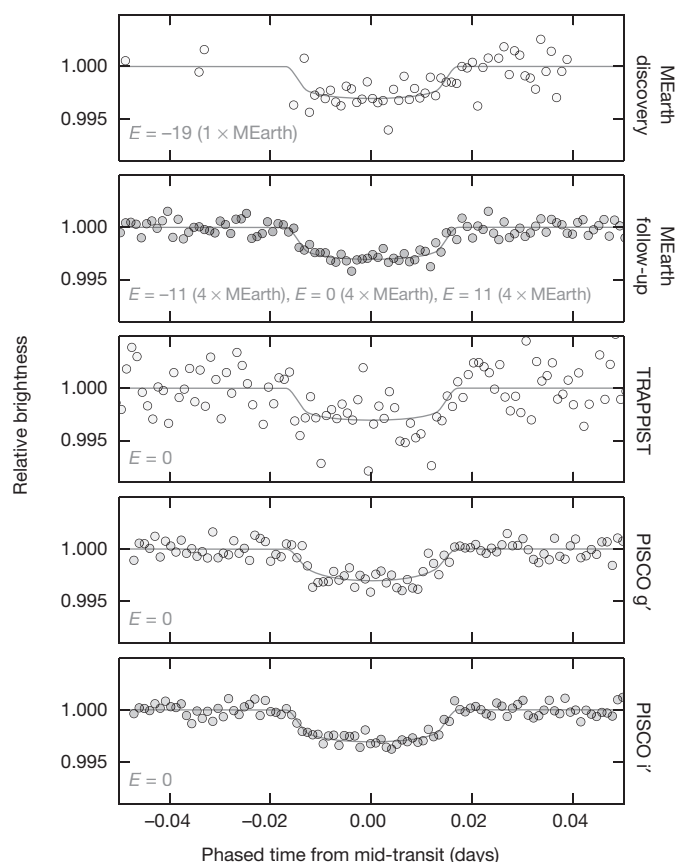
# A rocky planet transiting a nearby low-mass star

Zachory K. Berta-Thompson<sup>1,2</sup>, Jonathan Irwin<sup>2</sup>, David Charbonneau<sup>2</sup>, Elisabeth R. Newton<sup>2</sup>, Jason A. Dittmann<sup>2</sup>, Nicola Astudillo-Defru<sup>3</sup>, Xavier Bonfils<sup>4,5</sup>, Michaël Gillon<sup>6</sup>, Emmanuël Jehin<sup>6</sup>, Antony A. Stark<sup>2</sup>, Brian Stalder<sup>7</sup>, François Bouchy<sup>3,8</sup>, Xavier Delfosse<sup>4,5</sup>, Thierry Forveille<sup>4,5</sup>, Christophe Lovis<sup>3</sup>, Michel Mayor<sup>3</sup>, Vasco Neves<sup>9</sup>, Francesco Pepe<sup>3</sup>, Nuno C. Santos<sup>10,11</sup>, Stéphane Udry<sup>3</sup> & Anaël Wünsche<sup>4,5</sup>

**M-dwarf stars—hydrogen-burning stars that are smaller than 60 per cent of the size of the Sun—are the most common class of star in our Galaxy and outnumber Sun-like stars by a ratio of 12:1. Recent results have shown that M dwarfs host Earth-sized planets in great numbers<sup>1,2</sup>; the average number of M-dwarf planets that are between 0.5 to 1.5 times the size of Earth is at least 1.4 per star<sup>3</sup>. The nearest such planets known to transit their star are 39 parsecs away<sup>4</sup>, too distant for detailed follow-up observations to measure the planetary masses or to study their atmospheres. Here we report observations of GJ 1132b, a planet with a size of 1.2 Earth radii that is transiting a small star 12 parsecs away. Our Doppler mass measurement of GJ 1132b yields a density consistent with an Earth-like bulk composition, similar to the compositions of the six known exoplanets with masses less than six times that of the Earth and precisely measured densities<sup>5–11</sup>. Receiving 19 times more stellar radiation than the Earth, the planet is too hot to be habitable but is cool enough to support a substantial atmosphere, one that has probably been considerably depleted of hydrogen. Because the host star is nearby and only 21 per cent the radius of the Sun, existing and upcoming telescopes will be able to observe the composition and dynamics of the planetary atmosphere.**

We used the MEarth–South telescope array<sup>12</sup> to monitor the brightness of the star GJ 1132, starting on 28 January 2014. The array consists of eight 40-cm robotic telescopes located at the Cerro Tololo Inter-American Observatory (CTIO) in Chile, and observes a sample of M-dwarf stars that are within 33 parsecs of Earth and smaller than 0.35 solar radii. Since early 2014, the telescopes have gathered data almost every night that weather has permitted, following a strategy similar to that of the MEarth–North survey<sup>13</sup>. On 10 May 2015, GJ 1132 was observed at 25-minute cadence until a real-time analysis system identified a slight dimming of the star indicative of a possible ongoing transit, and commanded the telescope to observe the star continuously at 0.75-minute cadence. These triggered observations confirmed the presence of a transit with a sharp egress (Fig. 1). A search of extant data (4,208 observations over 333 nights) for periodic signals revealed a 1.6-day candidate that included this event and reached a detection statistic<sup>13</sup> of  $9.1\sigma$ . Follow-up photometry of subsequent predicted transits with four MEarth–South telescopes, the TRAPPIST (TRANSiting Planets and Planetesimals Small Telescope) telescope<sup>14</sup>, and the PISCO (Parallel Imager for Southern Cosmology Observations) multiband imager<sup>15</sup> on the Magellan Clay telescope confirmed the transit signal as being consistent with a planet-sized object that is blocking 0.26% of the star's light. We began precise Doppler monitoring with the HARPS (High Accuracy Radial Velocity Planet Searcher) spectrograph<sup>16</sup> on 6 June 2015 and gathered 25 radial velocity measurements for determining the planetary mass (Fig. 2).

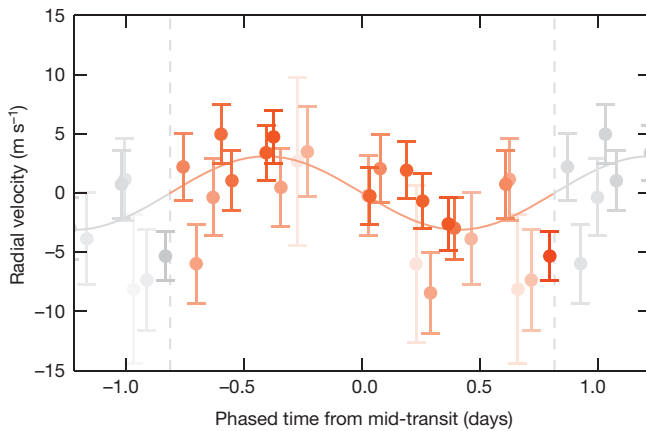
The distance to GJ 1132 has been measured through trigonometric parallax to be  $12.04 \pm 0.24$  parsecs<sup>17</sup>, a value that we independently validate with MEarth astrometry (see Methods). Together with empirical



**Figure 1 | Photometric measurements of transits of GJ 1132b.** Light curves from the MEarth–South, TRAPPIST and PISCO telescopes/imagers were fitted with a transit model (grey lines) and a Gaussian process noise model (subtracted from this plot), and averaged to 1.5-min bins for visual clarity. For MEarth–South, both the initial triggered ‘discovery’ observations and the subsequent ‘follow-up’ observations are shown. Labels indicate the transit event (with  $E$  as an integer number of planetary periods) and, for MEarth–South, the number of telescopes used. The opacities of binned points are inversely proportional to their assigned variances, representing their approximate weights in the model fit. The raw data and details of the fit are presented in Methods;  $g'$  and  $i'$  refer to the wavelength bandpasses used from the PISCO imager.

<sup>1</sup>Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. <sup>2</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. <sup>3</sup>Observatoire de Genève, Université de Genève, 51 chemin des Maillettes, 1290 Sauverny, Switzerland. <sup>4</sup>Université Grenoble Alpes, IPAG, F-38000 Grenoble, France. <sup>5</sup>CNRS, IPAG, F-38000 Grenoble, France. <sup>6</sup>Institut d'Astrophysique et de Géophysique, Université de Liège, Allée du 6 Août 17, Bâtiment B5C, 4000 Liège, Belgium. <sup>7</sup>Institute for Astronomy, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA. <sup>8</sup>Laboratoire d'Astrophysique de Marseille, UMR 6110 CNRS, Université de Provence, 38 rue Frédéric Joliot-Curie, 13388, Marseille Cedex 13, France. <sup>9</sup>Departamento de Física, Universidade Federal do Rio Grande do Norte, 59072-970 Natal, Rio Grande do Norte, Brazil. <sup>10</sup>Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal. <sup>11</sup>Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua Campo Alegre, 4169-007 Porto, Portugal.





**Figure 2 | Radial velocity changes over the orbit of GJ 1132b.**

Measurements of the star's line-of-sight velocity, taken by the HARPS spectrograph, are shown phased to the planetary orbital period determined from the light curves (orange points, with duplicates shown in grey). Error bars correspond to  $1\sigma$ . The darkness of each point is proportional to its weight in the model fit, which is the inverse of its variance as predicted by a radial velocity noise model. For a circular orbit, the star's reflex motion to the planet has a semi-amplitude of  $K_* = 2.76 \pm 0.92 \text{ m s}^{-1}$ .

relations among the intrinsic luminosities, masses and radii of M-dwarf stars<sup>18,19</sup>, the parallax enables us to estimate the mass and radius of GJ 1132. These estimates are not biased by physically associated luminous companions, which are ruled out by published photometry results and the HARPS spectra. Likewise, unassociated background stars are too faint in archival imaging at the current sky position of this high-proper-motion star to corrupt our estimates of the stellar parameters. Table 1 presents the physical properties of the star (GJ 1132) and planet (GJ 1132b), combining the inferred stellar properties with analyses of the transit light curves (Fig. 1) and radial velocity observations (Fig. 2). The radius of the planet is 40% that of GJ 1214b (ref. 20), a well studied mini-Neptune exoplanet that orbits with a similar period around a similar host star.

GJ 1132b's average density resembles that of the Earth, and is well matched by a rock/iron bulk composition. A theoretical mass–radius curve<sup>21</sup> for a two-layer planet composed of 75% magnesium silicate and 25% iron (by mass) is consistent with our estimates for GJ 1132b (Fig. 3). This model assumes that the core is pure iron, the mantle is pure magnesium silicate, and the interior contains no water<sup>21</sup>. These simplifications mean that the iron fraction should not be taken as absolute; the model simply represents a characteristic mass–radius locus that matches Earth and Venus. This same composition also matches the masses and radii of Kepler-78b (refs 8, 9), Kepler-10b (ref. 7), Kepler-93b (ref. 10), Kepler-36b (ref. 6), CoRoT-7b (ref. 5), and HD 219134b (ref. 11) to within  $1\sigma$ . All of these planets are smaller than 1.6 Earth radii, a transition radius above which most planets require thick hydrogen/helium envelopes to explain their densities<sup>22</sup>. At the  $1\sigma$  lower bound of GJ 1132b's estimated mass, models<sup>23</sup> indicate that replacing only 0.2% of the rock/iron mix with a hydrogen/helium layer would increase the planet's radius to 1.4 times that of the Earth, substantially larger than the observed value. Detection of GJ 1132b's mass is currently only at the  $3\sigma$  level, but continued Doppler monitoring will shrink the 35% mass uncertainty and enable more detailed comparison with other planets and compositional models.

We searched for additional planets both as other transits in the MEarth–South light curve and as periodic signals in the HARPS residuals. Although we made no notable discoveries, we highlight that compact, coplanar, multiple-planet systems are common around small stars<sup>24,25</sup>. Further exploration of the GJ 1132 system could reveal more, potentially transiting, planets.

As a relatively cool rocky exoplanet with an equilibrium temperature between 580 K (assuming a Bond albedo of 0) and 410 K (assuming

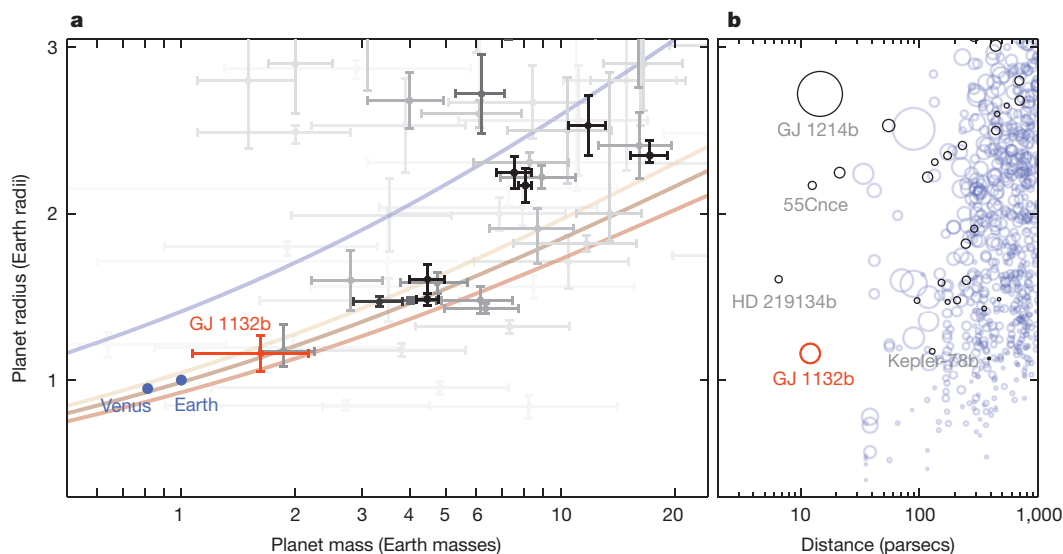
**Table 1 | System properties for GJ 1132b**

Parameter	Value
<b>Stellar parameters</b>	
Photometry	$V = 13.49, J = 9.245, K = 8.322$
Distance to star, $D_*$	$12.04 \pm 0.24$ parsecs
Mass of star, $M_*$	$0.181 \pm 0.019 M_\odot$
Density of star, $\rho_*$	$29.6 \pm 6.0 \text{ g cm}^{-3}$
Radius of star, $R_*$	$0.207 \pm 0.016 R_\odot$
Luminosity of star, $L_*$	$0.00438 \pm 0.00034 L_\odot$
Effective temperature, $T_{\text{eff}}$	$3,270 \pm 140 \text{ K}$
Metallicity, $[\text{Fe}/\text{H}]$	$-0.12 \pm 0.15$
Age of star, $\tau_*$	$> 5 \text{ Gyr}$
<b>Transit and radial velocity parameters</b>	
Orbital period, $P$ (days)	$1.628930 \pm 3.1 \times 10^{-5}$
Time of mid-transit, $t_0$ (BJD <sub>TDB</sub> ; days)	$2,457,184.55786 \pm 0.00032$
Eccentricity, $e$	0 (fixed)
Planet-to-star radius ratio, $R_p/R_*$	$0.0512 \pm 0.0025$
Scaled orbital distance, $a/R_*$	$16.0 \pm 1.1$
Impact parameter, $b$	$0.38 \pm 0.14$
Radial velocity semi-amplitude, $K_*$	$2.76 \pm 0.92 \text{ m s}^{-1}$
Systemic velocity, $\gamma_*$	$35 \pm 1 \text{ km s}^{-1}$
<b>Planet parameters</b>	
Radius of planet, $R_p$	$1.16 \pm 0.11 R_\oplus$
Mass of planet, $M_p$	$1.62 \pm 0.55 M_\oplus$
Density of planet, $\rho_p$	$6.0 \pm 2.5 \text{ g cm}^{-3}$
Surface gravity on planet, $g_p$	$1,170 \pm 430 \text{ cm s}^{-2}$
Escape velocity, $V_{\text{esc}}$	$13.0 \pm 2.3 \text{ km s}^{-1}$
Equilibrium temperature, $T_{\text{eq}}$	
assuming Bond albedo of 0.00	$579 \pm 15 \text{ K}$
assuming Bond albedo of 0.75	$409 \pm 11 \text{ K}$

Transit and radial velocity parameters were estimated from a Markov chain Monte Carlo (MCMC) analysis, including an external constraint on the stellar density when deriving  $P$ ,  $t_0$ ,  $R_p/R_*$ ,  $a/R_*$ , and  $b$  (see Methods). Planetary properties were derived from the combined stellar, transit, and radial velocity parameters.  $a$ , orbital semimajor axis; BJD<sub>TDB</sub>, barycentric Julian date in the barycentric dynamical time system;  $L_\odot$ , luminosity of the Sun;  $M_\odot$ , mass of the Sun;  $R_\odot$ , radius of the Sun;  $M_\oplus$ , mass of Earth;  $R_\oplus$ , radius of Earth.

a Venus-like Bond albedo of 0.75), GJ 1132b may have retained a substantial atmosphere. At these temperatures, the average thermal speeds of atoms or molecules heavier than helium are less than one-eighth of the escape velocity, suggesting an atmosphere potentially stable against thermal escape. This is not the case for the other rocky exoplanets for which precise densities are known, all of which are considerably hotter. The rocky planet Kepler-78b (refs 8, 9), which is comparable in size and density to GJ 1132b, receives 200 times more irradiation than GJ 1132b. Whether the atmosphere of GJ 1132b was initially dominated by hydrogen/helium-rich gas accreted from the primordial nebula or by volatiles outgassed from the planetary interior, its composition probably evolved substantially over the age of the system, which we estimate to exceed 5 billion years (gigayears, Gyr) (see Methods). Irradiated well beyond the runaway greenhouse limit<sup>26</sup>, surface water would extend up to high altitudes where it could be destroyed by photolysis and its hydrogen rapidly lost to space. When the star was young and bright at ultraviolet wavelengths, an atmosphere with high concentrations of water could lose hydrogen at the diffusion limit, of the order of  $10^{13}$  atoms per  $\text{cm}^2$  per second or 10 Earth oceans per gigayear. Depending on surface weathering processes, the oxygen left behind might persist as  $\text{O}_2$  in the atmosphere<sup>26,27</sup>. In this scenario, water would constitute a trace component in an atmosphere otherwise dominated by  $\text{O}_2$ ,  $\text{N}_2$  and  $\text{CO}_2$ . However, large uncertainties in the size of the initial hydrogen reservoir, in the history of the star's ultraviolet luminosity, in the contribution of late volatile delivery, and in the evolutionary effect of the system's likely spin–orbit synchronization preclude firm *a priori* statements about the composition of the atmosphere.

Future spectroscopic investigation of the planetary atmosphere will be enabled by the proximity and small radius of the star. When



**Figure 3 | Masses, radii, and distances of known transiting planets. a.** The radius and mass of GJ 1132b (orange) are shown, along with those of other exoplanets (grey). Also shown are mass-radius curves predicted by theoretical models<sup>21</sup> for planets composed of 100% H<sub>2</sub>O (blue line), and for two-component planets composed of MgSiO<sub>3</sub> on top of Fe cores that are 0% (light brown), 25% (darker brown) or 50% (red) of the total mass. Planets with smaller fractional mass and radius uncertainties are

darker. **b.** Symbol area is proportional to transit depth. In comparison with other transiting exoplanets, those with masses detected at  $>2.5\sigma$  (black) and those without masses detected at such level (blue), GJ 1132b is the most accessible terrestrial planet for spectroscopic observations of its atmosphere, owing to the proximity and small size of its parent star. All error bars in **a** are  $1\sigma$ .

viewed in transmission during transit, one scale height of an O<sub>2</sub>-rich atmosphere would overlap 10 parts per million (p.p.m.) of the stellar disk. For comparison, a 60-orbit Hubble Space Telescope transmission spectrum of GJ 1214b achieved a transit depth precision of 25 p.p.m. in narrow-wavelength bins<sup>28</sup>. Deeper Hubble observations of GJ 1132, which is 50% brighter than GJ 1214, would have the potential to detect molecular absorption features in GJ 1132b's atmosphere. Observations with the James Webb Space Telescope (JWST), set to launch in 2018, could measure the transmission spectrum over a broader wavelength range and require less telescope time. The long-wavelength capabilities of JWST may also allow it to detect the thermal emission from the planet; such emission represents 40–130 p.p.m. of the system flux at a wavelength of 10  $\mu$ m, and 160–300 p.p.m. of the flux at 25  $\mu$ m (for the range of albedos considered above). Provided that the planet is not too cloudy, combined transmission and emission spectra could ascertain the abundances of strongly absorbing molecular species. If such constraints on the dominant infrared opacity sources can be obtained, observations of the planet's thermal phase curve would be sensitive to complementary information, including the total atmospheric mass<sup>29,30</sup>. Such observations will inform our understanding of how the strong tides and intense stellar activity of the M-dwarf planetary environment influence the evolution of terrestrial atmospheres. This understanding will be important for the long-term goal of looking for life on planets orbiting nearby small stars.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 August; accepted 23 September 2015.**

1. Dressing, C. D. & Charbonneau, D. The occurrence rate of small planets around small stars. *Astrophys. J.* **767**, 95 (2013).
2. Morton, T. D. & Swift, J. The radius distribution of planets around cool stars. *Astrophys. J.* **791**, 10 (2014).
3. Dressing, C. D. & Charbonneau, D. The occurrence of potentially habitable planets orbiting M dwarfs estimated from the full Kepler dataset and an empirical measurement of the detection sensitivity. *Astrophys. J.* **807**, 45 (2015).
4. Muirhead, P. S. *et al.* Characterizing the cool KOIs. III. KOI 961: a small star with large proper motion and three small planets. *Astrophys. J.* **747**, 144 (2012).

5. Haywood, R. D. *et al.* Planets and stellar activity: hide and seek in the CoRoT-7 system. *Mon. Not. R. Astron. Soc.* **443**, 2517–2531 (2014).
6. Carter, J. A. *et al.* Kepler-36: a pair of planets with neighboring orbits and dissimilar densities. *Science* **337**, 556–559 (2012).
7. Dumusque, X. *et al.* The Kepler-10 planetary system revisited by Harps-N: a hot rocky world and a solid Neptune-mass planet. *Astrophys. J.* **789**, 154 (2014).
8. Pepe, F. *et al.* An Earth-sized planet with an Earth-like density. *Nature* **503**, 377–380 (2013).
9. Howard, A. W. *et al.* A rocky composition for an Earth-sized exoplanet. *Nature* **503**, 381–384 (2013).
10. Dressing, C. D. *et al.* The mass of Kepler-93b and the composition of terrestrial planets. *Astrophys. J.* **800**, 135 (2015).
11. Motalebi, F. *et al.* The HARPS-N rocky planet search I. HD 219134 b: a transiting rocky planet in a 4 planet system at 6.5 pc from the Sun. *Astron. Astrophys.* <http://dx.doi.org/10.1051/0004-6361/201526822> (2015).
12. Irwin, J. M. *et al.* The MEarth-North and MEarth-South transit surveys: searching for habitable super-Earth exoplanets around nearby M-dwarfs. In *18th Conference Cambridge Work on Cool Stars, Stellar Systems and the Sun* (eds van Belle, G. & Harris, H. C.) 767–772 (<http://adslabs.org/adsabs/abs/2015csss...18..767I/>) (2015).
13. Berta, Z. K., Irwin, J., Charbonneau, D., Burke, C. J. & Falco, E. E. Transit detection in the MEarth survey of nearby M dwarfs: bridging the clean-first, search-later divide. *Astron. J.* **144**, 145 (2012).
14. Gillon, M. *et al.* TRAPPIST: a robotic telescope dedicated to the study of planetary systems. *EPL Web Conf.* **11**, 06002 (2011).
15. Stalder, B. *et al.* PISCO: the Parallel Imager for Southern Cosmology Observations. In *Proc. SPIE* (eds Ramsay, S. K., McLean, I. S. & Takami, H.) Vol. 9147, 91473Y (2014).
16. Mayor, M. *et al.* Setting new standards with HARPS. *Messenger* **114**, 20–24 (2003).
17. Jao, W.-C. *et al.* The solar neighborhood XIII: parallax results from the CTIOPI 0.9-m program—stars with  $\mu \geq 1''/\text{year}$  (MOTION sample). *Astron. J.* **129**, 1954 (2005).
18. Delfosse, X. *et al.* Accurate masses of very low mass stars: IV. Improved mass-luminosity relations. *Astron. Astrophys.* **364**, 217–224 (2000).
19. Hartman, J. D. *et al.* HATS-6b: a warm Saturn transiting an early M dwarf star, and a set of empirical relations for characterizing K and M dwarf planet hosts. *Astron. J.* **149**, 166 (2015).
20. Charbonneau, D. *et al.* A super-Earth transiting a nearby low-mass star. *Nature* **462**, 891–894 (2009).
21. Zeng, L. & Sasselov, D. D. A detailed model grid for solid planets from 0.1 through 100 Earth masses. *Publ. Astron. Soc. Pacif.* **125**, 227–239 (2013).
22. Rogers, L. A. Most 1.6 Earth-radius planets are not rocky. *Astrophys. J.* **801**, 41 (2015).
23. Lopez, E. D. & Fortney, J. J. Understanding the mass-radius relation for sub-Neptunes: radius as a proxy for composition. *Astrophys. J.* **792**, 1 (2014).
24. Ballard, S. & Johnson, J. A. The Kepler dichotomy among the M dwarfs: half of systems contain five or more coplanar planets. Preprint at <http://adslabs.org/adsabs/abs/2014arXiv1410.4192B/> (2014).

25. Muirhead, P. S. *et al.* Kepler-445, Kepler-446 and the occurrence of compact multiples orbiting mid-M dwarf stars. *Astrophys. J.* **801**, 18 (2015).
26. Kasting, J. F., Whitmire, D. P. & Reynolds, R. T. Habitable zones around main sequence stars. *Icarus* **101**, 108–128 (1993).
27. Luger, R. & Barnes, R. Extreme water loss and abiotic O<sub>2</sub> buildup on planets throughout the habitable zones of M dwarfs. *Astrobiology* **15**, 119–143 (2015).
28. Kreidberg, L. *et al.* Clouds in the atmosphere of the super-Earth exoplanet GJ 1214b. *Nature* **505**, 69–72 (2014).
29. Selsis, F., Wordsworth, R. & Forget, F. Thermal phase curves of nontransiting terrestrial exoplanets 1. Characterizing atmospheres. *Astron. Astrophys.* **532**, A1 (2011).
30. Koll, D. D. B. & Abbot, D. S. Deciphering thermal phase curves of dry, tidally locked terrestrial planets. *Astrophys. J.* **802**, 21 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the staff at the Cerro Tololo Inter-American Observatory for assistance in the construction and operation of MEarth-South; J. Winn and J. Berta-Thompson for comments on the manuscript; S. Seager and A. Zsom for conversations that improved the work; L. Delrez for her independent analysis of the TRAPPIST data; and J. Eastman, D. Dragomir and R. Siverd for their efforts to observe additional transits. The MEarth Project acknowledges funding from the David and Lucile Packard Fellowship for Science and Engineering, and the National Science Foundation, and a grant from the John Templeton Foundation. The opinions expressed here are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. The development of the PISCO imager was supported by the National Science Foundation. HARPS observations were made with European Southern Observatory (ESO) Telescopes at the La Silla Paranal Observatory.

TRAPPIST is a project funded by the Belgian Fund for Scientific Research, with the participation of the Swiss National Science Foundation. Z.K.B.-T. is funded by the MIT Torres Fellowship for Exoplanet Research. X.B., X.D., T.F. and A.W. acknowledge the support of the French Agence Nationale de la Recherche and the European Research Council. M.G. and E.J. are FNRS Research Associates. V.N. acknowledges a CNPq/BJT Post-Doctorate fellowship and partial financial support from the INCT INEspaço. N.C.S. acknowledges the support from the Portuguese National Science Foundation (FCT) as well as the COMPETE program.

**Author Contributions** The MEarth team (D.C., J.I., Z.K.B.-T., E.R.N. and J.A.D.) discovered the planet, organized the follow-up observations, and led the analysis and interpretation. Z.K.B.-T. analysed the light curve and radial velocity data, and wrote the manuscript. J.I. designed, installed, maintains, and operates the MEarth-South telescope array, identified the first triggered transit event, and substantially contributed to the analysis and interpretation. D.C. leads the MEarth Project, and assisted in analysis and writing the manuscript. E.R.N. determined the metallicity, kinematics, and rotation period of the star. J.A.D. confirmed the star's trigonometric parallax and helped install the MEarth-South telescopes. The HARPS team (N.A.-D., X.B., F.B., X.D., T.F., C.L., M.M., V.N., F.P., N.C.S., S.U. and A.W.) obtained spectra for Doppler velocimetry, with N.A.-D. and X.B. leading the analysis of those data. M.G. and E.J. gathered photometric observations with TRAPPIST. A.A.S. and B.S. gathered photometric observations with PISCO. All authors read and discussed the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.K.B.-T. ([zkbt@mit.edu](mailto:zkbt@mit.edu)).



## METHODS

**Distance to the star.** GJ 1132's coordinates are 10:14:51.77–47:09:24.1 (International Celestial Reference System, epoch 2000.0), with proper motions of  $(-1.046; 416)$  milliarcseconds (mas) per year and a trigonometric parallax of  $\pi = 83.07 \pm 1.69$  mas, as determined by the RECONS (Research Consortium on Nearby Stars) survey<sup>17</sup>. We qualitatively confirm this parallax with independent observations from MEarth–South, using analyses like those that have been applied to the northern survey<sup>31</sup>. The motion of GJ 1132 relative to background stars in MEarth–South imaging closely matches the prediction made by the RECONS parallax (Extended Data Fig. 1). We do not quote the value of  $\pi$  derived from MEarth–South because we have not yet cross-validated the astrometric performance of the system against other measurements. Literature photometric observations of GJ 1132 include photoelectric photometry ( $U = 16.51 \pm 0.03$ ,  $B = 15.17 \pm 0.03$ )<sup>32</sup>, charge-coupled device (CCD) photometry ( $V = 13.49 \pm 0.03$ ,  $R_C = 12.26 \pm 0.02$ ,  $I_C = 10.69 \pm 0.02$ )<sup>17</sup>, 2MASS near-infrared photometry ( $J = 9.245 \pm 0.026$ ,  $H = 8.666 \pm 0.027$ ,  $K_s = 8.322 \pm 0.027$ )<sup>33</sup>, and WISE infrared photometry ( $W1 = 8.170 \pm 0.023$ ,  $W2 = 8.000 \pm 0.020$ ,  $W3 = 7.862 \pm 0.018$ ,  $W4 = 7.916 \pm 0.184$ ). The colour ( $V-K_s = 5.168 \pm 0.040$ ) and absolute magnitude ( $M_V = 13.088 \pm 0.054$ ) of GJ 1132 are consistent with those of single M4V dwarfs<sup>34</sup>.

**Metallicity of the star.** Before discovering the planet, we gathered a near-infrared spectrum of GJ 1132 with the FIRE spectrograph on the Magellan Baade telescope. We shifted the spectrum to a zero-velocity wavelength scale<sup>35</sup>, measured equivalent widths and compared the spectra by eye with solar metallicity spectral type standards<sup>35</sup>. The spectrum indicates a near-infrared spectral type of M4V–M5V (Extended Data Fig. 2), slightly later than the optical spectral type of M3.5V listed in the PMSU (Palomar/Michigan State University) catalogue<sup>36</sup>. Using the measured equivalent width of the K-band sodium feature (4.7 Å) and an empirical calibration<sup>35</sup> that has been corrected for its known temperature dependence<sup>37</sup>, we estimate the stellar metallicity to be  $[\text{Fe}/\text{H}] = -0.12 \pm 0.15$  and quote this value in Table 1. For comparison, a relation using additional spectral regions and calibrated for stars of GJ 1132's spectral type and earlier<sup>38</sup> also yields  $[\text{Fe}/\text{H}] = -0.1$ , while one for GJ 1132's spectral type and later<sup>39</sup> yields  $[\text{Fe}/\text{H}] = -0.2$  (both with uncertainties of about 0.15 dex).

**Mass of the star.** Dynamical mass measurements of M-dwarfs show that tight relationships exist between near-infrared absolute magnitudes and stellar mass<sup>18</sup>. We use these calibrations to calculate masses from the  $J$ ,  $H$  and  $K$  magnitudes (after converting between the 2MASS and CIT photometric systems). Taking the mean of these masses and adopting an uncertainty that is the quadrature sum of the 2.7% error propagated from the measurement uncertainties and the 10% scatter we assume for the relations, we adopt a stellar mass of  $M_* = 0.181 \pm 0.019 M_\odot$ , where  $M_*$  is the mass of the star and  $M_\odot$  is the mass of the Sun.

**Radius of the star.** From this mass, we use an empirical  $M_* - \rho_*$  relation<sup>19</sup> calibrated to eclipsing binary systems to estimate a density of  $\rho_* = 29.6 \pm 6.0 \text{ g cm}^{-3}$ , corresponding to  $R_* = 0.207 \pm 0.016 R_\odot$  for GJ 1132. We adopt those values, noting that they agree with two other mass–radius relations: the radius predicted by long-baseline optical interferometry of single stars<sup>40</sup> is  $R_* = 0.211 \pm 0.014 R_\odot$ , and that by the Dartmouth evolutionary models<sup>41</sup> is  $R_* = 0.200 \pm 0.016 R_\odot$  (for  $[\text{Fe}/\text{H}] = -0.1$ , assuming a uniform prior on age between 1 Gyr and 10 Gyr). The quoted errors do not include an assumed intrinsic scatter in any of the mass–radius relations, but the consistency among the three estimates suggests that any contribution from scatter would be smaller than the uncertainty propagated from the stellar mass.

**Bolometric luminosity of the star.** We combine the parallax and photometry with bolometric corrections to determine the total luminosity of GJ 1132, testing three different relations to estimate bolometric corrections from colour. The Mann *et al.* relation<sup>42</sup> between  $\text{BC}_V$  and  $V-J$  colour yields a bolometric luminosity of  $0.00402 L_\odot$ . The Leggett *et al.* relation<sup>43</sup> between  $\text{BC}_K$  and  $I-K$  colour yields  $0.00442 L_\odot$ . The Pecaut and Mamajek compilation of literature bolometric corrections<sup>44</sup>, when interpolated in  $V-K_s$  colour to determine  $\text{BC}_V$ , yields  $0.00469 L_\odot$ . We adopt the mean of these three values, with an uncertainty that is the quadrature sum of the systematic error (the 6.3% standard deviation of the different estimates) and the uncertainty propagated from the measurement uncertainties (about 5% in all three cases), as our final estimate of the bolometric luminosity:  $L_* = 0.00438 \pm 0.00034 L_\odot$ . From this, we calculate the stellar effective temperature as  $T_{\text{eff}} = 5,772 \text{ K} \times (L_*/L_\odot)^{1/4} \times (R_*/R_\odot)^{-1/2} = 3,270 \pm 140 \text{ K}$ . The luminosity and temperature we infer<sup>37</sup> from the FIRE spectra ( $L_* = 0.0044 \pm 0.001 L_\odot$ ,  $T_{\text{eff}} = 3,130 \pm 120 \text{ K}$ ) are consistent with the quoted values.

**Age of the star.** GJ 1132's motion through the Galaxy of  $(U, V, W) = (-39, -31, -2) \text{ km s}^{-1}$  is consistent with a kinematically older stellar population. M4 dwarfs tend to show strong H $\alpha$  emission for about 4 Gyr<sup>45</sup>; the lack of H $\alpha$  emission in the HARPS spectrum indicates that GJ 1132 is probably older than that. The star instead shows weak H $\alpha$  absorption, which is an indicator of non-zero magnetic

activity in stars as cool as this<sup>46</sup>. We detect emission in the Ca II H line, with a weak intensity that is comparable to that of Barnard's Star and other slowly rotating stars in the HARPS M-dwarf sample<sup>47</sup>. Applying published methods<sup>48</sup> to the MEarth–South photometry, we measure a rotation period of 125 days for GJ 1132 (Extended Data Fig. 3). M-dwarfs spin more slowly as they age, with less massive stars reaching longer rotation periods at old ages<sup>48</sup>. Therefore, the rotation period suggests that the system could be almost as old as Barnard's Star ( $0.16 M_\odot$ ), which has a 130-day rotation period<sup>49</sup> and is 7–13 Gyr old<sup>50</sup>. GJ 1132's age is probably comparable to or greater than that of Proxima Centauri ( $0.12 M_\odot$ ), which has an 83-day period<sup>49,51</sup> and is 5–7 Gyr old, assuming it to be coeval with the gyro-chronologically and asteroseismically age-dated a Centauri system<sup>52</sup>. From these comparisons, we conclude that GJ 1132 is probably older than 5 Gyr.

**Assumption of a circular orbit.** Calculating the timescale for tidal circularization<sup>53</sup> as  $t_{\text{circ}} = \frac{2PQ}{63\pi} \times \frac{M_p}{M_*} \times \left(\frac{a}{R_p}\right)^5$ , assuming a tidal quality factor of  $Q = 100$  appropriate

for rocky exoplanets, and where  $P$  is the orbital period and  $a$  is the orbital semi-major axis, yields  $4 \times 10^5$  years for GJ 1132b. This is much shorter than the age of the system; therefore, we assume the eccentricity to be negligible and fix it to 0 in our estimation of the other planetary properties. Perturbations from other (undetected) planets in the system could potentially induce a small equilibrium eccentricity.

**Photometric observations and analysis.** The high-cadence light curves analysed here were gathered over four nights using six telescopes on three mountains. The triggered event (event  $E = -19$ , relative to the ephemeris in Table 1) was collected on MEarth–South telescope 3. At the time of this triggered detection, 296 stars in the MEarth–South sample had been observed at least 100 times, but only nine of these stars had been observed as thoroughly as GJ 1132. After identification of the periodic signal, we obtained follow-up photometry with MEarth–South telescopes 2, 3, 4 and 8 on the nights of 23 May 2015, 10 June 2015 and 28 June 2015 ( $E = -11, 0$  and  $11$ ). Exposure times were 18 s, yielding a 47-s cadence for the high-cadence light curves. We observed the transit on the night of 10 June 2015 ( $E = 0$ ) with the TRAPPIST telescope, in a wide  $I+z$  bandpass, with exposure times of 10 s and a cadence of 21 s. We also observed this transit with the PISCO multiband imager, which was installed on the Magellan Clay telescope and undergoing a commissioning run, taking 10-s exposures simultaneously in  $g'$ ,  $r'$ ,  $i'$ ,  $z'$  bandpasses at a cadence of 36 s. We actively defocused the Clay telescope, resulting in donut-shaped point spread functions spreading each star's light over about 200 pixels. The PISCO  $r'$  and  $z'$  light curves exhibited unexplained systematics and were not used for further analysis.

To the high-cadence light curves, we fitted a transit model<sup>54</sup> with the following parameters: orbital period  $P$ , time of mid-transit  $t_0$ , planet-to-star radius ratio  $R_p/R_*$ , star-to-orbit radius ratio  $R_*/a$ , impact parameter  $b$ , and separate baseline flux levels for each transit observed on each telescope. The PISCO and TRAPPIST photometry showed trends that correlated with airmass, so we also included coefficients for a trend linear with airmass as free parameters. Limb-darkening was treated with a quadratic approximation, using fixed coefficients<sup>55</sup> calculated from a PHOENIX atmosphere model for a 3,300 K,  $[\text{Fe}/\text{H}] = 0$ ,  $\log g = 5$  star. These coefficients ( $u_1, u_2$ ) were: (0.1956, 0.3700) for MEarth and TRAPPIST (both approximated as  $1/3i' + 2/3z'$ ), (0.4790, 0.3491) for PISCO  $g'$  (approximated as Sloan  $g'$ ), and (0.2613, 0.3737) for PISCO  $i'$  (approximated as Sloan  $i'$ ). We performed a Levenberg–Marquardt maximization of the posterior probability of this model, rejected  $>4\sigma$  outliers among the data, repeated the maximization, and increased the per-point uncertainty estimates until each transit exhibited a reduced  $\chi^2 \leq 1$ . This resulted in typical per-point uncertainties of 2.6 (MEarth), 3.6 (TRAPPIST), 1.9 (PISCO  $g'$ ), and 1.2 (PISCO  $i'$ ) mmag. Most light curves showed some evidence for time-correlated noise in their residuals. To marginalize over the uncertainty introduced by these correlations, we use the Gaussian process regression package George<sup>56</sup> to model each set of transit residuals as a Gaussian process with non-zero covariance between the datapoints<sup>57</sup>. We use a Matérn-3/2 kernel function to describe this covariance as a function of separation in time, and include two free parameters per transit:  $\log t_{\text{gp}}$  (where  $t_{\text{gp}}$  is the correlation timescale) and  $\log a_{\text{gp}}$  (where  $a_{\text{gp}}$  is an amplitude). We mapped the probability distribution of the model parameters using the emcee implementation<sup>58</sup> of an affine-invariant Markov chain Monte Carlo (MCMC) sampler<sup>59</sup>, assuming flat priors on all parameters. Extended Data Fig. 4 shows raw light curves of individual transits, along with model curves sampled from the posterior.

The ingress time ( $\tau = 3.7 \pm 1.0$  min from first to second contact) is measured imprecisely in this fit, compared to the total transit duration ( $T_{\text{tot}} = 47.0 \pm 1.4$  min from first to fourth contact). Systematic astrophysical errors (incorrect limb-darkening, starspots, unidentified transit timing variations) can bias light-curve estimates of the ingress time, and therefore also the parameters  $R_*/a$  and  $b$  that depend strongly on  $\tau$ . Therefore, we estimate the planet properties without relying on this ingress measurement, by including external constraints on both the

stellar mass and the stellar density. We sample values of  $M_*$  and  $\rho_*$  from Gaussian distributions ( $0.181 \pm 0.019 M_\odot$  and  $29.6 \pm 6.0 \text{ g cm}^{-3}$ ), use Kepler's Third Law to compute  $R_*/a$  from the stellar density and the period, and then calculate  $b$  from the transit duration and these new  $R_*/a$  samples. We quote the marginalized values and uncertainties for these and other light-curve parameters in Table 1. This procedure forces consistency between the light-curve fit and the inferred stellar properties and shifts the estimates from  $R_*/a = 0.0738 \pm 0.0092$  and  $b = 0.58 \pm 0.14$  (inferred from the light curves alone) to the values in Table 1. The effect on  $P$ ,  $t_0$ , or  $R_p/R_*$  is negligible. If we ignore the external  $\rho_*$  information, the inferred stellar and planetary radii would be consistent with those in Table 1 but with larger error bars ( $R_* = 0.243 \pm 0.031 R_\odot$  and  $R_p = 1.37 \pm 0.22 R_\oplus$ ).

**Radial velocity observations and analysis.** We first gathered reconnaissance spectra of GJ 1132 with the CHIRON<sup>60</sup> spectrograph on the SMARTS 1.5-metre telescope at CTIO. Once these spectra ruled out large radial-velocity variations corresponding to binary star companions, we began observations with the HARPS spectrograph on the La Silla 3.6-metre telescope. We gathered 25 HARPS spectra of GJ 1132 between 6 June 2015 and 18 July 2015, using an exposure time of 40 min. The spectra span 380 nm to 680 nm in wavelength. For each observation, we constructed a comparison template by co-adding all the other spectra and measured the relative radial velocity as the shift required to minimize the  $\chi^2$  of the difference between the spectrum and this template<sup>61,62</sup>. Telluric lines were masked using a template of lines made with a different, much larger, data set. We see no evidence for rotational broadening of the spectra, indicating that the projected rotational velocity,  $v \sin i$ , is less than  $2 \text{ km s}^{-1}$ . The median of the internal radial velocity uncertainty estimates<sup>61</sup> is  $3 \text{ m s}^{-1}$  per observation.

Over the timespan of our observations, these radial velocities are dominated by the orbital motion of the known transiting planet, rather than by other planets, blended stellar binaries, or stellar activity. The tallest peak in the periodogram of the radial velocities is consistent with the transit-derived period for GJ 1132b. The line bisectors show no significant correlation with the measured velocities (Pearson correlation  $P$ -value of 0.27) and no clear periodicities. The H $\alpha$  equivalent width and the full-width half-maximum of the HARPS cross-correlation function are not correlated with the velocities (Pearson correlation  $P$ -values of 0.72 and 0.57).

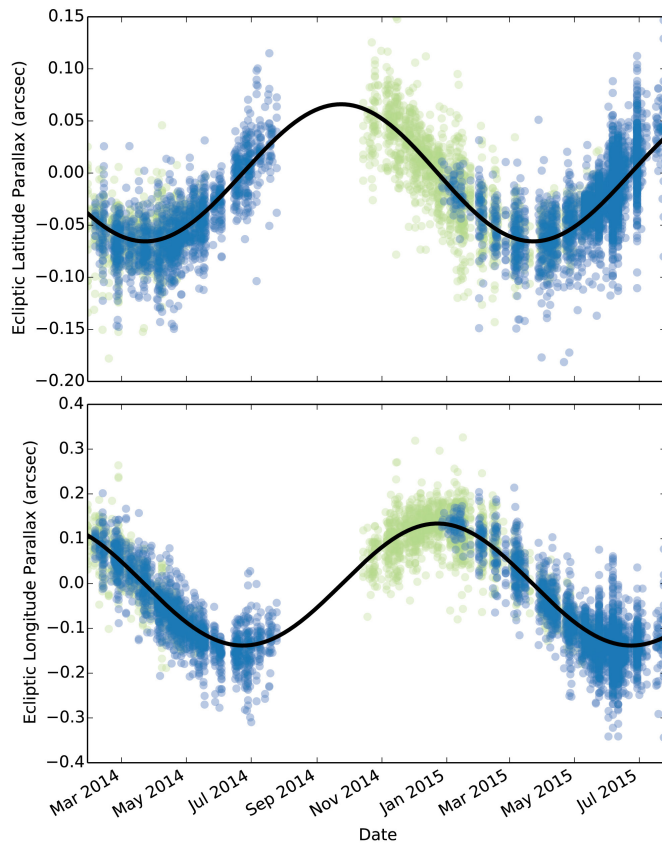
We fit the velocities with a model corresponding to a circular orbit for the planet, with flat priors on three free-floating parameters: the radial velocity semi-amplitude  $K_*$ , the systemic velocity  $\gamma_*$ , and a stellar jitter term  $\sigma_{*,\text{jitter}}$  that is added in quadrature to the internal velocity uncertainties. We include all terms that depend on  $\sigma_{*,\text{jitter}}$  in the likelihood, as well as the usual  $\chi^2$  term.  $P$  and  $t_0$  were fixed to the values determined from the transit analysis. We use emcee<sup>58</sup> to sample from the posterior probability distributions of these parameters, and quote marginalized uncertainties on  $K_*$  in Table 1. In this fit, the value of  $K_*$  is  $\geq 0$  in 99.7% of the MCMC samples. The inferred value of  $\sigma_{*,\text{jitter}}$  is smaller than the individual uncertainties ( $< 1.9 \text{ m s}^{-1}$  at 68% confidence), and a good fit is obtained if  $\sigma_{*,\text{jitter}}$  is fixed to 0 ( $\chi^2 = 26.67$  for 23 degrees of freedom). The  $\chi^2$  of a fit where both  $K_*$  and  $\sigma_{*,\text{jitter}}$  are fixed to 0 is significantly worse (38.56 for 24 degrees of freedom). As a check, we repeated the MCMC fit allowing the phase ( $t_0$ ) to float. This radial velocity fit predicted the known mid-transit times to within  $1.5\sigma$  ( $3.0 \pm 1.9 \text{ h}$  after the true transit times). Relaxing the assumption of a circular orbit yields inferred distributions of  $(e \cos \omega, e \sin \omega)$  that are consistent with (0, 0) but substantially increases the uncertainty on the semi-amplitude ( $K_* = 3.6 \pm 2.4 \text{ m s}^{-1}$ ). With future Doppler measurements, it will be possible to measure the planet's eccentricity and independently test the assumption we make here that tides damped the eccentricity to small values.

**Transiting exoplanet population comparison.** Figure 3 compares GJ 1132b to other known transiting exoplanets. Data for this plot were drawn from the NASA Exoplanet Archive<sup>63</sup> on 25 August 2015. Egregious errors in the catalogue were replaced with appropriate literature values. In the mass–radius panel, individual planets are shaded with a white-to-black greyscale that is inversely proportional to  $(\sigma_M/M_p)^2 + (\sigma_R/R_p)^2$ , where  $\sigma_M$  and  $\sigma_R$  are the uncertainties on the planet mass,  $M_p$ , and radius,  $R_p$ . Planets with precise measurements are dark, and those with large mass or radius uncertainties are light. In the distance–radius panel, some planets did not have distances listed in the Exoplanet Archive. For those systems, we calculated approximate distances from the J-band magnitudes, estimated stellar radii, effective temperatures, and a table of bolometric corrections<sup>44</sup> (interpolating in  $T_{\text{eff}}$  to estimate  $BC_J$ ).

**Code availability.** Analyses were conducted primarily in Python. Although not cleanly packaged for general use, for the sake of transparency we make the custom code used for transit and radial velocity fitting available at <http://github.com/>

zkbt/transit. It relies on three freely available packages: eb (<http://github.com/mdwarfgeek/eb>), emcee (<http://github.com/dfm/emcee>), and George (<http://github.com/dfm/george>). Code used to generate the exoplanet population comparison is available at <http://github.com/zkbt/exopop>.

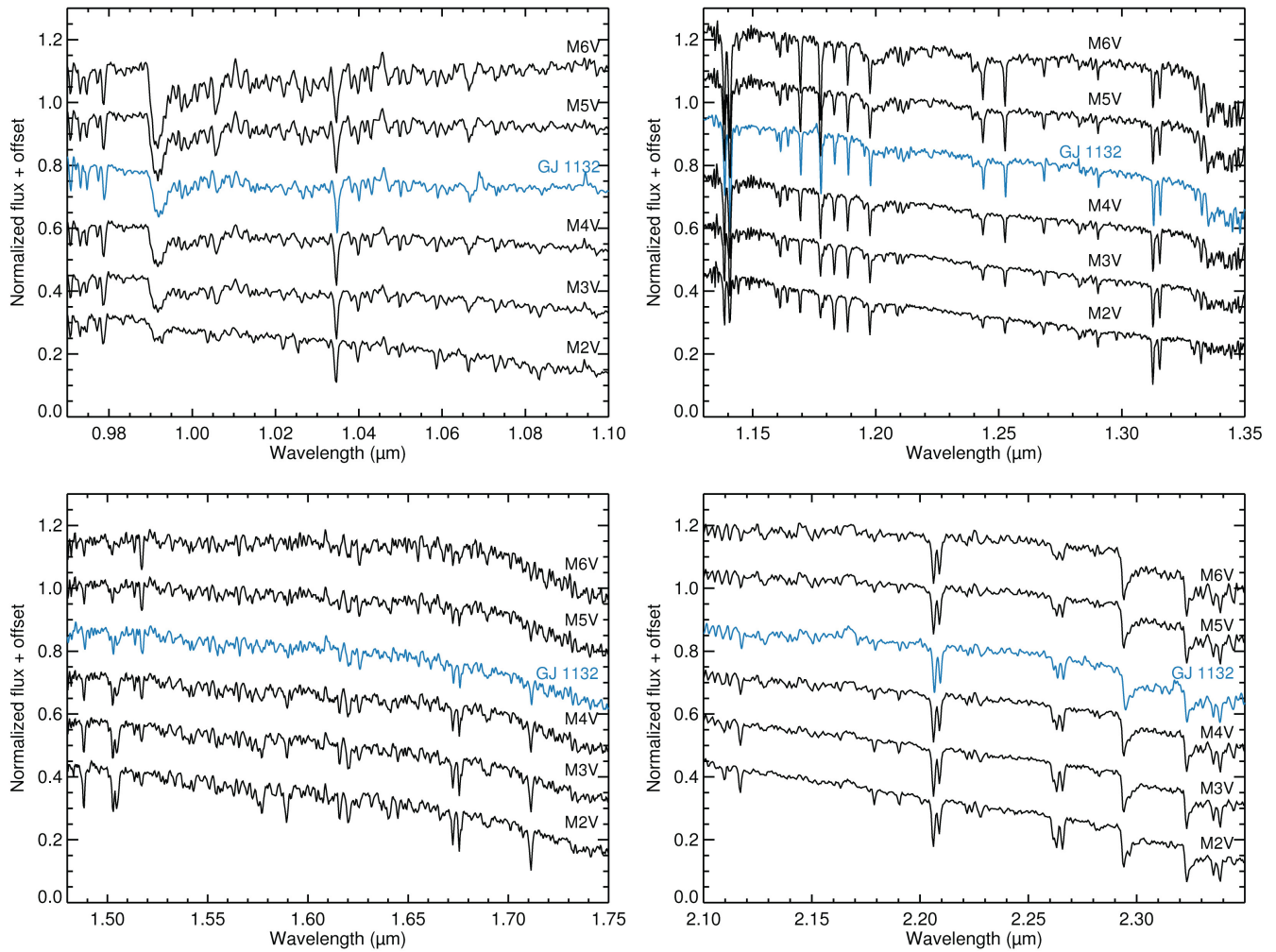
31. Dittmann, J. A., Irwin, J. M., Charbonneau, D. & Berta-Thompson, Z. K. Trigonometric parallaxes for 1507 nearby mid-to-late M dwarfs. *Astrophys. J.* **784**, 156 (2014).
32. Eggen, O. J. Catalogs of proper-motion stars. I. Stars brighter than visual magnitude 15 and with annual proper motion of 1 arcsec or more. *Astrophys. J.* **39** (Suppl.), 89 (1979).
33. Skrutskie, M. F. *et al.* The two micron all sky survey (2MASS). *Astron. J.* **131**, 1163–1183 (2006).
34. Winters, J. G. *et al.* The solar neighborhood. XXXV. Distances to 1404 M dwarf systems within 25 pc in the southern sky. *Astron. J.* **149**, 5 (2014).
35. Newton, E. R. *et al.* Near-infrared metallicities, radial velocities, and spectral types for 447 nearby M dwarfs. *Astron. J.* **147**, 20 (2014).
36. Hawley, S. L., Gizis, J. E. & Reid, N. I. The Palomar/MSU nearby star spectroscopic survey. II. The southern M dwarfs and investigation of magnetic activity. *Astron. J.* **113**, 1458 (1997).
37. Newton, E. R., Charbonneau, D., Irwin, J. & Mann, A. W. An empirical calibration to estimate cool dwarf fundamental parameters from H-band spectra. *Astrophys. J.* **800**, 85 (2015).
38. Mann, A. W., Brewer, J. M., Gaidos, E., Lépine, S. & Hilton, E. J. Prospecting in late-type dwarfs: a calibration of infrared and visible spectroscopic metallicities of late K and M dwarfs spanning 1.5 dex. *Astron. J.* **145**, 52 (2013).
39. Mann, A. W. *et al.* Prospecting in ultracool dwarfs: measuring the metallicities of mid- and late-M dwarfs. *Astron. J.* **147**, 160 (2014).
40. Boyajian, T. S. *et al.* Stellar diameters and temperatures. II. Main-sequence K- and M-stars. *Astrophys. J.* **757**, 112 (2012).
41. Dotter, A. *et al.* The Dartmouth stellar evolution database. *Astrophys. J.* **178** (Suppl.), 89–101 (2008).
42. Mann, A. W., Feiden, G. A., Gaidos, E., Boyajian, T. & von Braun, K. How to constrain your M dwarf: measuring effective temperature, bolometric luminosity, mass, and radius. *Astrophys. J.* **804**, 64 (2015).
43. Leggett, S. K., Allard, F., Geballe, T. R., Hauschildt, P. H. & Schweitzer, A. Infrared spectra and spectral energy distributions of late M and L dwarfs. *Astrophys. J.* **548**, 908–918 (2001).
44. Pecaut, M. J. & Mamajek, E. E. Intrinsic colors, temperatures, and bolometric corrections of pre-main-sequence stars. *Astrophys. J.* **208** (Suppl.), 9 (2013).
45. West, A. A. *et al.* Constraining the age-activity relation for cool stars: the Sloan digital sky survey data release 5 low-mass star spectroscopic sample. *Astron. J.* **135**, 785–795 (2008).
46. Walkowicz, L. M. & Hawley, S. L. Tracers of chromospheric structure. I. Observations of Ca II K and H $\alpha$  in M dwarfs. *Astron. J.* **137**, 3297–3313 (2009).
47. Bonfils, X. *et al.* The HARPS search for southern extra-solar planets. XXXI. The M-dwarf sample. *Astron. Astrophys.* **549**, A109 (2013).
48. Irwin, J. *et al.* On the angular momentum evolution of fully-convective stars: rotation periods for field M-dwarfs from the MEarth transit survey. *Astrophys. J.* **727**, 56 (2010).
49. Benedict, G. F. *et al.* Photometry of Proxima Centauri and Barnard's Star using Hubble space telescope fine guidance sensor 3: a search for periodic variations. *Astron. J.* **116**, 429–439 (1998).
50. Feltzing, S. & Bensby, T. The galactic stellar disc. *Phys. Scr.* **2008**, T133 (2008).
51. Kiraga, M. & Stepien, K. Age-rotation-activity relations for M dwarf stars. *Acta Astron.* **57**, 149–172 (2007).
52. Mamajek, E. E. & Hillenbrand, L. A. Improved age estimation for solar-type dwarfs using activity-rotation diagnostics. *Astrophys. J.* **687**, 1264 (2008).
53. Goldreich, P. & Soter, S. Q in the solar system. *Icarus* **5**, 375–389 (1966).
54. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580**, L171–L175 (2002).
55. Claret, A., Hauschildt, P. H. & Witte, S. New limb-darkening coefficients for PHOENIX/1D model atmospheres. *Astron. Astrophys.* **546**, A14 (2012).
56. Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W. & O'Neil, M. Fast direct methods for Gaussian processes and the analysis of NASA Kepler mission data. Preprint at <http://arxiv.org/abs/1403.6015> (2014).
57. Gibson, N. P. *et al.* A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy. *Mon. Not. R. Astron. Soc.* **419**, 2683–2694 (2012).
58. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. Emcee: the MCMC hammer. *Publ. Astron. Soc. Pacif.* **125**, 306–312 (2013).
59. Goodman, J. & Weare, J. Ensemble samplers with affine invariance. *Comm. App. Math. Comp. Sci.* **5**, 65–80 (2010).
60. Tokovinin, A. *et al.* CHIRON—a fiber fed spectrometer for precise radial velocities. *Publ. Astron. Soc. Pacif.* **125**, 1336–1347 (2013).
61. Bouchy, F., Pepe, F. & Queloz, D. Fundamental photon noise limit to radial velocity measurements. *Astron. Astrophys.* **374**, 733–739 (2001).
62. Astudillo-Defru, N. *et al.* The HARPS search for southern extra-solar planets. *Astron. Astrophys.* **575**, A119 (2015).
63. Akeson, R. L. *et al.* The NASA exoplanet archive: data and tools for exoplanet research. *Publ. Astron. Soc. Pacif.* **125**, 989–999 (2013).



**Extended Data Figure 1 | Astrometry of GJ 1132 from MEarth-South.**

Measurements of the star GJ 1132's position in MEarth-South images, along the directions of ecliptic latitude (top) and longitude (bottom). As described elsewhere<sup>31</sup>, a fitted offset between data gathered at a field rotation of 0° (blue) and 180° (green) has been removed. The published RECONS proper motion<sup>17</sup> has been subtracted, and a model fixed to the published 83.07 mas parallax (black line) closely matches the MEarth-South observations.

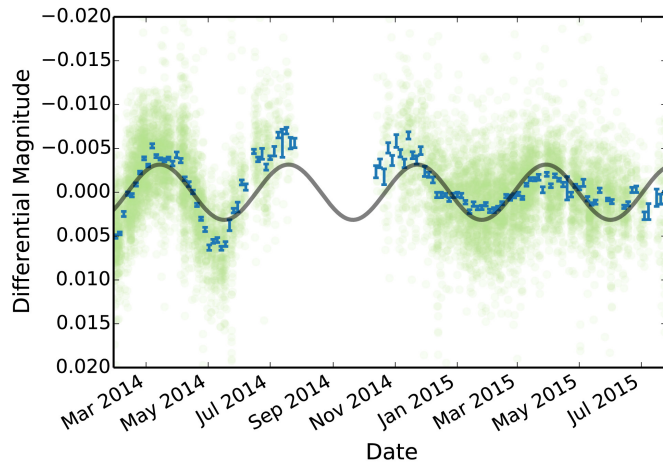




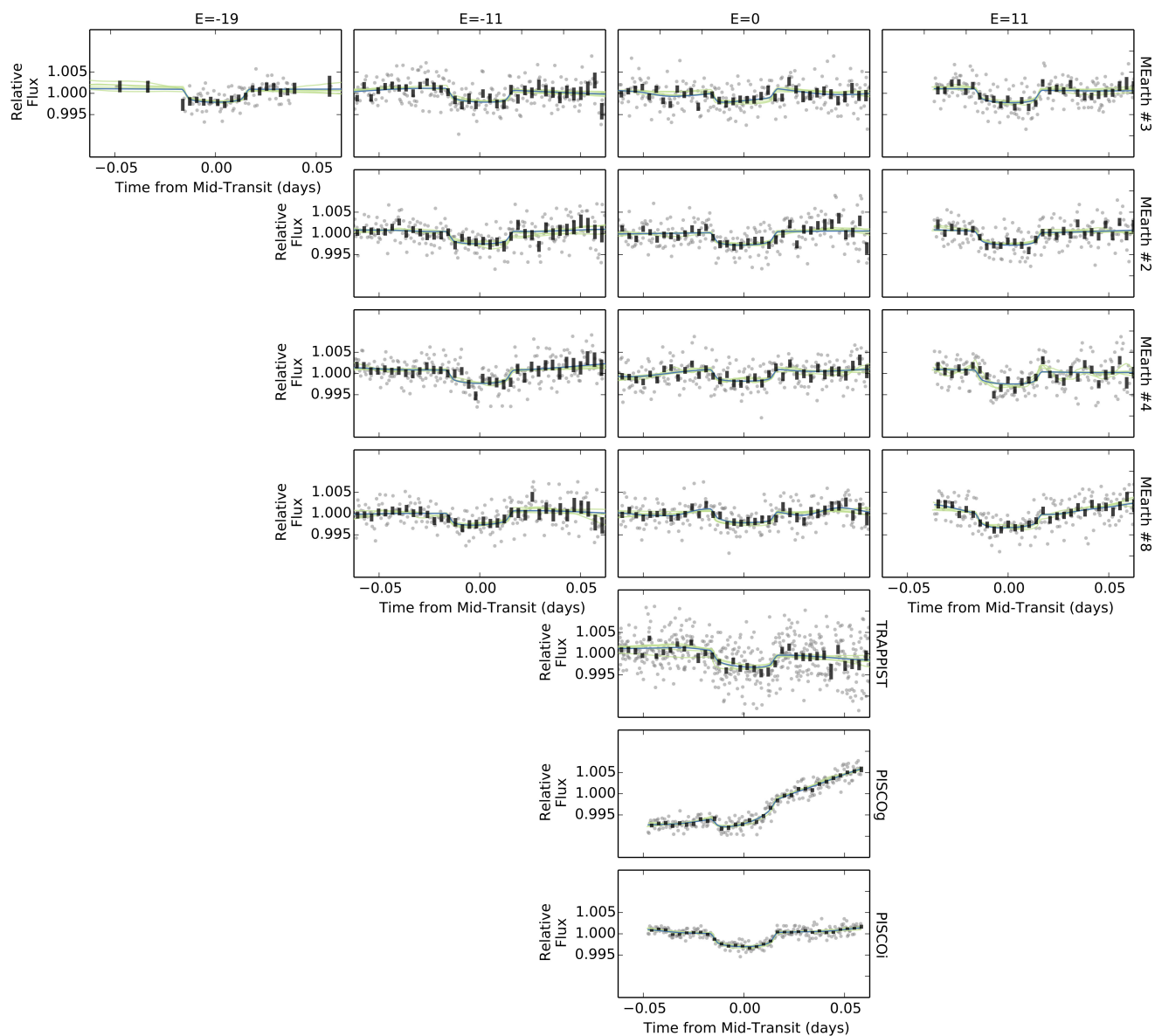
### Extended Data Figure 2 | Near-infrared spectrum of GJ 1132.

Observations of GJ 1132's spectrum obtained with the FIRE spectrograph on the Magellan Baade telescope are compared in the  $z$  (top left),  $J$  (top right),  $H$  (bottom left) and  $K$  (bottom right) telluric windows to the

solar-metallicity composite spectral type standards from ref. 35. The FIRE spectra have been smoothed to match the  $R=2,000$  resolution of the standards. GJ 1132's near-infrared spectral type is M4V–M5V.



**Extended Data Figure 3 | Photometric starspot modulations of GJ 1132.** MEarth–South photometry (with dots representing single pointings and error bars representing  $\pm 1\sigma$  uncertainty ranges on weighted averages over four-day bins) probes starspots that are rotating in and out of view, and indicates that GJ 1132 has a rotation period of approximately 125 days. The rotational modulation was identified using a methodology similar to that used in previous MEarth work<sup>48</sup>.



**Extended Data Figure 4 | Raw transit light curves of GJ 1132b.** Light curves are shown both unbinned (grey points) and in five-minute bins (black bars, representing the  $\pm 1\sigma$  uncertainty range for the weighted average in each bin), and separated by telescope (row) and transit event

(column). Model curves are shown, with the Gaussian process noise model conditioned on the observations, for parameters sampled from the posterior (green) and for the maximum likelihood parameters (blue). This is the complete set of light-curve data behind the transit parameter fits.



# Entangling two transportable neutral atoms via local spin exchange

A. M. Kaufman<sup>1,2</sup>, B. J. Lester<sup>1,2</sup>, M. Foss-Feig<sup>3</sup>, M. L. Wall<sup>1</sup>, A. M. Rey<sup>1,2</sup> & C. A. Regal<sup>1,2</sup>

To advance quantum information science, physical systems are sought that meet the stringent requirements for creating and preserving quantum entanglement. In atomic physics, robust two-qubit entanglement is typically achieved by strong, long-range interactions in the form of either Coulomb interactions between ions or dipolar interactions between Rydberg atoms<sup>1–4</sup>. Although such interactions allow fast quantum gates, the interacting atoms must overcome the associated coupling to the environment and cross-talk among qubits<sup>5–8</sup>. Local interactions, such as those requiring substantial wavefunction overlap, can alleviate these detrimental effects; however, such interactions present a new challenge: to distribute entanglement, qubits must be transported, merged for interaction, and then isolated for storage and subsequent operations. Here we show how, using a mobile optical tweezer, it is possible to prepare and locally entangle two ultracold neutral atoms, and then separate them while preserving their entanglement<sup>9–11</sup>. Ground-state neutral atom experiments have measured dynamics consistent with spin entanglement<sup>10,12,13</sup>, and have detected entanglement with macroscopic observables<sup>14,15</sup>; we are now able to demonstrate position-resolved two-particle coherence via application of a local gradient and parity measurements<sup>1</sup>. This new entanglement-verification protocol could be applied to arbitrary spin-entangled states of spatially separated atoms<sup>16,17</sup>. The local entangling operation is achieved via spin-exchange interactions<sup>9–11</sup>, and quantum tunnelling is used to combine and separate atoms. These techniques provide a framework for dynamically entangling remote qubits via local operations within a large-scale quantum register.

Internal spin states of particles provide robust and long-lived storage for quantum information. Although engineered spin-dependent interactions can realize entangling gates between spins, they also predispose the system to strong environmental coupling and decoherence<sup>9</sup>. Spin-exchange interactions, which arise from a combination of quantum statistics and spin-independent forces, afford a promising alternative route to entanglement generation, and have been explored with both electrons (in quantum dots) and atoms<sup>9–11,18–20</sup>. When two particles interact, their interaction energy depends on the spatial symmetry of the two-particle wavefunction. If the particles have spin but are otherwise identical, the symmetry of the two-particle spin state directly determines the spatial symmetry of their wavefunction: for repulsive interactions, two bosons (fermions) of opposite spin in a triplet configuration  $T$  experience enhanced (suppressed) interactions, while the converse occurs for the singlet spin state  $S$ ; see Fig. 1b. By preparing a superposition of the triplet and singlet, dynamical quantum beats result in the exchange of spin between the particles. In our experiment, we prepare atoms of opposing spin in the lowest two motional states ( $e$  and  $g$ ) of an optical tweezer potential that we represent as  $|\uparrow_e, \downarrow_g\rangle$ , which results in equal population of the singlet and triplet spatial wavefunctions  $\psi_S(\mathbf{r}_1, \mathbf{r}_2)$  and  $\psi_T(\mathbf{r}_1, \mathbf{r}_2)$ , where  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are the spatial coordinates associated with particles 1 and 2, respectively. The difference in the contact interaction energy between these states yields spin-exchange dynamics at a rate  $J_{\text{ex}}$ , which depends on the  $s$ -wave

scattering length and two-particle density (see Methods). Specifically, it gives rise to the dynamics

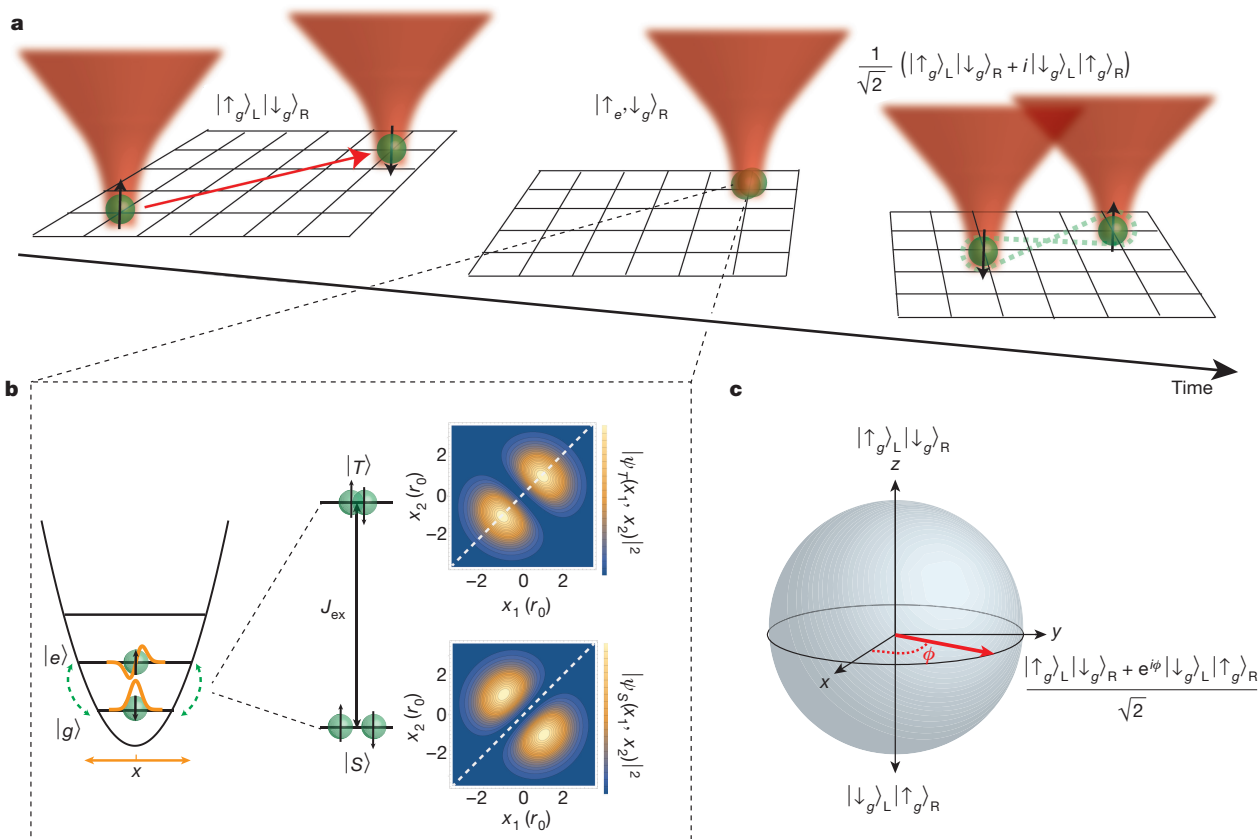
$$|\psi(t)\rangle = |\uparrow_e, \downarrow_g\rangle \cos(J_{\text{ex}}t/2\hbar) + i|\downarrow_e, \uparrow_g\rangle \sin(J_{\text{ex}}t/2\hbar) \quad (1)$$

This evolution is associated with the effective spin-dependent Hamiltonian  $H = J_{\text{ex}}\mathbf{S}_e \cdot \mathbf{S}_g$ , where  $\mathbf{S}_e$  and  $\mathbf{S}_g$  are the spin operators for the respective motional states. A spin-entangled state  $(|\uparrow_e, \downarrow_g\rangle + i|\downarrow_e, \uparrow_g\rangle)/\sqrt{2}$  can be created by allowing the state to evolve for an exchange time of  $\pi\hbar/2J_{\text{ex}}$ .

We schematically represent the experiment in Fig. 1a, in which separated optical tweezers on the left (L) and right (R) each containing a single atom are dynamically reconfigured to produce spin-exchange dynamics and non-local entanglement. Atoms are combined in the right optical tweezer where spin exchange creates the entangled state  $(|\uparrow_e, \downarrow_g\rangle_R + i|\downarrow_e, \uparrow_g\rangle_R)/\sqrt{2}$ . Importantly, in our experiments we convert this entanglement into spatial spin correlations by separating the atoms into two tweezers (Fig. 1a) to yield a state  $(|\uparrow_g\rangle_L|\downarrow_g\rangle_R + i|\downarrow_g\rangle_L|\uparrow_g\rangle_R)/\sqrt{2}$ . Although verification of such entanglement is a standard tool in ion and Rydberg experiments<sup>1,3,4,21</sup>, spatially resolved detection of the entanglement present in interacting systems of ground-state neutral atoms is challenging. Theoretical proposals have studied ways of detecting spatial<sup>22</sup> and spin<sup>16</sup> entanglement, and very recently experimental progress has been made using a quantum gas microscope<sup>17</sup>. We devised a protocol that yields rotations of a two-qubit entangled state on the associated Bloch sphere via a combination of a magnetic-field gradient and global spin rotations (Fig. 1c), allowing detection of basis-independent correlations for arbitrary entangled states of the form  $\frac{1}{\sqrt{2}}(|\uparrow_g\rangle_L|\downarrow_g\rangle_R + e^{i\phi}|\downarrow_g\rangle_L|\uparrow_g\rangle_R)$ . Our protocol is applicable to qubit pairs in a large quantum register<sup>11</sup>, to interacting spins in a Bose–Hubbard chain<sup>13,16</sup>, and to strongly interacting fermions featuring antiferromagnetic correlations<sup>23–25</sup>.

The experiment begins by loading two thermal <sup>87</sup>Rb atoms into two separate optical tweezer potentials<sup>26</sup>. Each atom is then separately laser-cooled to the 3D ground state via Raman-sideband cooling, leaving a 3D ground-state fraction of 90(7)% (refs 5, 27, 28); the uncertainty in parentheses here and below indicates the standard error. The atoms are initialized in opposite spin states with a fidelity in the range of 80%–91% via single-atom addressing (see Methods). We then prepare the pure two-particle state  $|\uparrow_e, \downarrow_g\rangle_R$  on a single tweezer by reducing the separation between the two optical tweezers, and using our capability to control the tunnel coupling between wavelength-scale optical tweezer traps<sup>24,28</sup>. As illustrated in Fig. 2a, the spacing between the two tweezers is decreased until tunnelling is appreciable; however, unlike our previous work<sup>28</sup>, we dynamically shift to an asymmetric configuration such that the ground state of one optical tweezer is near-resonant with the first (radial) excited state of the other optical tweezer. We then perform adiabatic passage by slowly tuning the relative tweezer depths linearly in time, which coherently transfers the left atom into the right optical tweezer over the 12-ms duration of the ramp (see Methods).

<sup>1</sup>JILA, National Institute of Standards and Technology and University of Colorado, Boulder, Colorado 80309, USA. <sup>2</sup>Department of Physics, University of Colorado, Boulder, Colorado 80309, USA. <sup>3</sup>Joint Quantum Institute and the National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA.



**Figure 1 | Experimental overview.** **a**, Left, two particles (green spheres), trapped in separate optical tweezers (red), are initiated in their motional ground state and in opposing spin states (black arrows). Dynamical reconfiguration of optical tweezer traps merges the particles into the same optical tweezer in specific motional states (centre), and spin exchange entangles the particles. The atoms are separated into different optical tweezers, and the entanglement between the particles is experimentally detected (right). **b**, Exchange interaction. Left, two atoms of opposite spin are prepared in the ground ( $|g\rangle$ ) and excited ( $|e\rangle$ ) motional states of the  $x$  direction. For bosonic atoms the singlet channel anti-symmetrizes in the motional degree of freedom, yielding

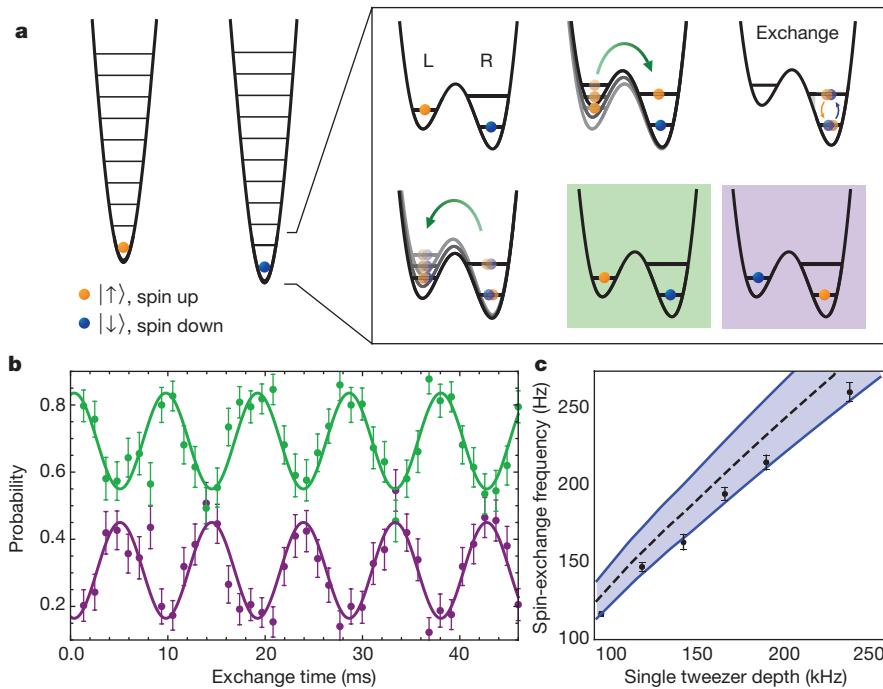
$\psi_S(\mathbf{r}_1, \mathbf{r}_2) = (\psi_g(\mathbf{r}_1)\psi_e(\mathbf{r}_2) - \psi_e(\mathbf{r}_1)\psi_g(\mathbf{r}_2))/\sqrt{2}$ . Right, from one-dimensional spatial wavefunctions we see that the anti-symmetrization prevents the two particles at positions  $x_1$  and  $x_2$  (in units of the oscillator length  $r_0$ ) from occupying the same position (dashed white line), which is true for all choices of  $y, z$ . This leads to zero interaction energy for the singlet channel and  $J_{\text{ex}}$  for the triplet channel with associated wavefunction  $\psi_T(\mathbf{r}_1, \mathbf{r}_2) = (\psi_g(\mathbf{r}_1)\psi_e(\mathbf{r}_2) + \psi_e(\mathbf{r}_1)\psi_g(\mathbf{r}_2))/\sqrt{2}$ . **c**, Studying entanglement of the separated particles. On the Bloch sphere (grey), the exchange interactions ideally result in a state pointing along the  $\pm y$  axis, but, in general, the state with associated Bloch vector (red arrow) can point in any direction in the equatorial plane with associated coherence angle  $\phi$ .

After a desired evolution time in the presence of exchange dynamics, the adiabatic passage is applied in reverse, yielding a motional-state mapping of the excited-state atom back into the ground state of the left well. We can then read out the spin populations to verify the exchange oscillations by ejecting atoms in the  $|\uparrow\rangle$  state from the optical tweezers, and imaging the atom occupancy in each of the two tweezers. With this procedure, we can ascertain what the spin and motional degrees of freedom of each of the two atoms were when they occupied the same optical tweezer.

Exchange oscillations in our experiment are shown in Fig. 2b, and show the expected anti-correlated behaviour. To display these data, we exclude from our analysis experiments in which imperfections in our spin preparation lead to the spins remaining aligned; such events yield a static contribution to the signal. We observe undamped oscillations out to times as long as 100 ms, despite the fact that the single-particle coherence time between  $|\uparrow\rangle$  and  $|\downarrow\rangle$  is less than 1 ms, owing to magnetic-field fluctuations. This is an expected feature of the entangled states created by the exchange interaction: homogeneous magnetic-field fluctuations induce a global phase on the two-particle superposition and, as such, leave quantum measurements unaffected. Hence, the state occupies a so-called ‘decoherence-free subspace’<sup>10,29</sup>. We can also control the frequency of the spin oscillations by modifying the depth of the optical tweezer in which the exchange occurs, which tunes the two-particle density. To study this

dependence, we prepare  $|\uparrow_e, \downarrow_g\rangle_R$  and linearly increase the depth of the tweezer in 5 ms, allow evolution of the exchange, and then ramp back in reverse and perform the second adiabatic passage. We can model the 3D non-separable potential of our optical tweezer trap<sup>30</sup>, and find agreement (Fig. 2c) between the calculated and measured spin-exchange frequency  $J_{\text{ex}}/(2\pi\hbar)$  (see Methods).

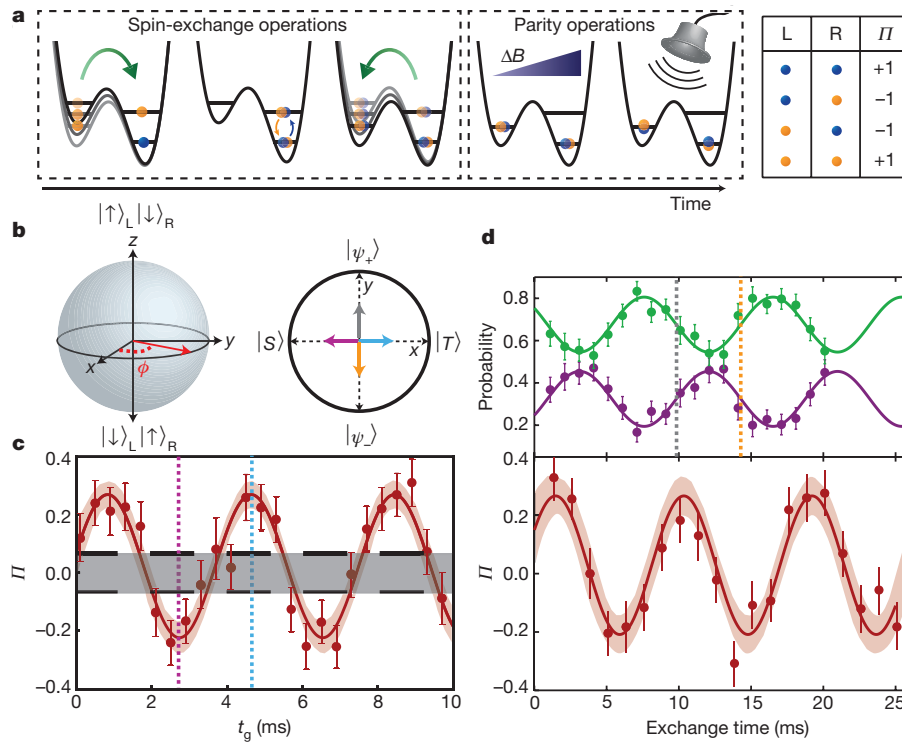
The measurements presented thus far have shown correlations in single-particle spin states. However, to ensure that future operations can retain and propagate quantum information, one must verify that the phase coherence within the entangled state is preserved upon separating the particles. The entanglement verification protocol for separated atoms is summarized in Fig. 3a. For explanatory purposes, we first focus on the case when the particles are separated after an exchange time of  $t_{\text{ent}} = n\pi\hbar/2J_{\text{ex}}$  where  $n$  is an odd integer. The entangled state after the second adiabatic passage is  $|\psi_{\pm}\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_L|\downarrow\rangle_R \pm i|\downarrow\rangle_L|\uparrow\rangle_R)$ , omitting from now on the ground-state ( $|g\rangle$ ) motional subscripts to simplify notation. The  $|\psi_{\pm}\rangle$  states correspond to the grey and orange Bloch vectors, respectively, in Fig. 3b. We then apply a magnetic-field gradient that imposes a difference,  $\hbar\delta$ , in the  $|\uparrow\rangle \leftrightarrow |\downarrow\rangle$  single-atom-transition energy between the left and right optical tweezer. By applying the gradient for a time  $t_g$ , a transformation  $|\psi_{\pm}\rangle \rightarrow \frac{1}{\sqrt{2}}(|\uparrow\rangle_L|\downarrow\rangle_R \pm ie^{i\delta t_g}|\downarrow\rangle_L|\uparrow\rangle_R)$  is achieved. As a function of  $t_g$ , the state rotates between the singlet (pink in



**Figure 2 | Direct observation of spin-exchange dynamics between two atoms.** **a**, Preparation of motional-state configuration and detection in the double-well potential (see key for properties of the atoms). Initially, an atom is prepared in the ground state of each of two spatially separated optical tweezers (left panel). The atoms are combined in one of the optical tweezers via tunnelling (green arrows), allowed to interact for a controlled amount of time during which spin exchange occurs, and then separated into two optical tweezers for detection (right panel). Green and purple backgrounds show the colour coding for states shown in **b**. **b**, Using post-selection on our spin preparation, we plot the likelihood to measure the state  $|\uparrow_g\rangle_L|\downarrow_g\rangle_R$  (green) and  $|\downarrow_g\rangle_L|\uparrow_g\rangle_R$  (purple) as a function of time between the end of the first adiabatic passage and the beginning of the second. **c**, Measured spin-exchange oscillation frequency as a function of optical tweezer depth. The dashed black line is the predicted exchange frequency from a parameter-free model of the potential (see Methods); the shaded region indicates the effect of systematic uncertainties on this prediction. All error bars are the standard error in the measurement.

Fig. 3b) and triplet (blue) with angular frequency  $\delta$ . We then apply a global  $\pi/2$  pulse in the  $\{|\uparrow\rangle, |\downarrow\rangle\}$  subspace. This pulse maps the singlet back to itself, while it maps the triplet to a Bell state  $\frac{1}{\sqrt{2}}(|\uparrow\rangle_L|\uparrow\rangle_R + |\downarrow\rangle_L|\downarrow\rangle_R)$ . Therefore, by measuring the probability that the spins are aligned or anti-aligned as a function of  $t_g$ , we can observe singlet–triplet oscillations whose amplitude characterizes the

two-particle coherence. We quantify this probability with the parity  $\Pi(t_g) = \sum_j P_j (-1)^j$ , where  $P_j$  is the likelihood to measure  $j$  atoms in the spin-down state<sup>1–4</sup>. The parity is equivalently the projection of the Bloch vector in Fig. 3b onto the  $x$  axis before the  $\pi/2$  pulse, and hence the gradient is essential because, although entangled, the states  $|\psi_{\pm}\rangle$  (grey, orange) exhibit zero parity after application of a  $\pi/2$  pulse.



**Figure 3 | Detection of non-local entanglement.**

**a**, Procedure for creating and detecting entanglement. The spin-exchange procedure of Fig. 2a is applied to entangle the atoms (left panel), and then, upon separating them, a magnetic-field gradient and microwaves are introduced (middle panel). The underlying entanglement is revealed by the parity of the final spin configurations of the particles (right panel). See Methods for details. **b**, Four different Bloch vector orientations in the experiment. Right, the grey and orange orientations correspond to the outcome of the spin-exchange dynamics (after the second adiabatic passage), which yield rotations in the  $y$ – $z$  plane of the Bloch sphere (left). The pink and blue orientations correspond to the points of peak parity, and are accessed by applying a magnetic-field gradient that rotates the states about the  $z$  axis. **c**, After creating  $|\psi_{\pm}\rangle$ , we plot the measured parity  $\Pi$  as a function of gradient time  $t_g$ . The grey bar is the bound on parity oscillation contrast delineating separable and entangled states, which accounts for imperfect spin preparation. The coloured dashed lines are at times when the associated states indicated in **b** are created. **d**, Measured parity as the exchange time is varied and correspondingly the atoms are entangled and unentangled. In the lower plot, we set  $t_g$  such that it rotates  $|\psi_{\pm}\rangle$  to  $|T\rangle$  and then measure the parity. The upper plot is the same experiment without the parity detection, that is, the protocol of Fig. 2. The dashed lines indicate times when the corresponding states indicated in **b** are produced. In **c** and **d**, the error bars in the data plots are the standard error, and the pink swaths show the 95% confidence bands.



We demonstrate the outcome of the verification protocol on the state  $|\psi_+\rangle$  in Fig. 3c. We plot  $\Pi(t_g)$  after the microwave  $\pi/2$  pulse, and observe oscillations in the parity signal as the gradient time  $t_g$  is scanned. The contrast of these oscillations is consistent with what is expected given the exchange oscillation contrast in Fig. 2, and non-vanishing parity oscillation would certify entanglement in the ideal case of perfect spin preparation. However, we have imperfect spin preparation, and the erroneous spin populations outside the  $\{|\downarrow\downarrow\rangle_L|\uparrow\uparrow\rangle_R, |\uparrow\uparrow\rangle_L|\downarrow\downarrow\rangle_R\}$  manifold could lead to parity oscillations even in the absence of entanglement. We have derived a condition on the parity oscillation contrast that is necessary and sufficient to certify entanglement, and is the simplest way to see there is entanglement in our system (for a full derivation of this condition and its relation to other entanglement metrics, see Methods). We relate the measured parity contrast,  $C$ , to the measured probabilities ( $P^{\uparrow\uparrow}, P^{\downarrow\downarrow}$ ) that the spins are erroneously prepared in the same spin-state: if  $C > C_{\text{bnd}} = 4(P^{\uparrow\uparrow}P^{\downarrow\downarrow})^{1/2}$ , then the state is entangled. By directly measuring the spin populations (see Methods) and their associated uncertainty, we ascertain  $C_{\text{bnd}} = 0.133(25)$  as indicated by the dashed lines in Fig. 3c. The observed parity oscillation contrast  $C = 0.49(4)$  exceeds  $C_{\text{bnd}}$  by more than  $7\sigma$ , certifying the presence of entanglement in the final state of the separated spins. We verify entanglement without correcting the measured parity for experimental imperfections, such as single-atom loss due to background collisions.

Whereas in Fig. 3c we varied the parity detection parameters via  $t_g$ , in Fig. 3d we measure the dependence of the parity on the exchange time at fixed  $t_g$ , thereby observing oscillations as the exchange interactions periodically entangle and unentangle the two atoms. We fix  $t_g$  in the parity detection such that the entangled state  $|\psi_+\rangle$  (grey lines in Fig. 3b, d) is rotated to a peak in  $\Pi$ , corresponding to the creation of the triplet (blue lines in Fig. 3b, c). Because this  $t_g$  amounts to a  $\pi/2$  rotation about the  $z$  axis of the Bloch sphere, it will also rotate  $|\psi_-\rangle$  to the singlet, which corresponds to maximal negative parity. In the lower panel of Fig. 3d, we show how the parity measured under these conditions oscillates at the exchange frequency  $J_{\text{ex}}/(2\pi\hbar)$ . For comparison, in the upper panel, we show the measured exchange oscillations (purple, green) without the parity detection. At the linear points of the exchange oscillations, one expects maximal entanglement corresponding to states  $|\psi_+\rangle$  (grey) and  $|\psi_-\rangle$  (orange) and thus the extremal points of the parity. At the minima and maxima of the exchange oscillations, the atoms are unentangled and the parity vanishes.

We have demonstrated the entanglement of remote qubits using spin-exchange interactions and a general protocol for detecting spin entanglement in a variety of systems. In a large register, tuning the entanglement phase could be achieved by a far-detuned focused probe that changes the local effective magnetic field experienced by a single qubit, and the qubits can be arranged to allow the passage of the optical tweezers without perturbing the qubits<sup>11,31</sup>. Although in this work we focus on quantum information applications, the ability to control spin and motion of individual neutral atoms will allow the study of intriguing microscopic models in condensed-matter physics, such as the Kondo lattice model<sup>32</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 July; accepted 30 September 2015.

Published online 2 November 2015.

1. Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
2. Leibfried, D. *et al.* Experimental demonstration of a robust, high-fidelity geometric two ion-qubit phase gate. *Nature* **422**, 412–415 (2003).
3. Wilk, T. *et al.* Entanglement of two individual neutral atoms using Rydberg blockade. *Phys. Rev. Lett.* **104**, 010502 (2010).

4. Isenhower, L. *et al.* Demonstration of a neutral atom controlled-NOT quantum gate. *Phys. Rev. Lett.* **104**, 010503 (2010).
5. Monroe, C. *et al.* Resolved-sideband Raman cooling of a bound atom to the 3D zero-point energy. *Phys. Rev. Lett.* **75**, 4011–4014 (1995).
6. Blakestad, R. *et al.* High-fidelity transport of trapped-ion qubits through an X-junction trap array. *Phys. Rev. Lett.* **102**, 153002 (2009).
7. Home, J. P. *et al.* Complete methods set for scalable ion trap quantum information processing. *Science* **325**, 1227–1230 (2009).
8. Béguin, L., Vernier, A., Chicireanu, R., Lahaye, T. & Browaeys, A. Direct measurement of the van der Waals interaction between two Rydberg atoms. *Phys. Rev. Lett.* **110**, 263201 (2013).
9. Hayes, D., Julienne, P. & Deutsch, I. Quantum logic via the exchange blockade in ultracold collisions. *Phys. Rev. Lett.* **98**, 070501 (2007).
10. Anderlini, M. *et al.* Controlled exchange interaction between pairs of neutral atoms in an optical lattice. *Nature* **448**, 452–456 (2007).
11. Weitenberg, C., Kuhr, S., Mølmer, K. & Sherson, J. Quantum computation architecture using optical tweezers. *Phys. Rev. A* **84**, 032322 (2011).
12. Mandel, O. *et al.* Controlled collisions for multi-particle entanglement of optically trapped atoms. *Nature* **425**, 937–940 (2003).
13. Fukuhara, T. *et al.* Quantum dynamics of a mobile spin impurity. *Nature Phys.* **9**, 235–241 (2013).
14. Lücke, B. *et al.* Twin matter waves for interferometry beyond the classical limit. *Science* **334**, 773–776 (2011).
15. Strobel, H. *et al.* Fisher information and entanglement of non-Gaussian spin states. *Science* **345**, 424–427 (2014).
16. Mazza, L., Rossini, D., Fazio, R. & Endres, M. Detecting two-site spin-entanglement in many-body systems with local particle-number fluctuations. *New J. Phys.* **17**, 013015 (2015).
17. Fukuhara, T. *et al.* Spatially resolved detection of a spin-entanglement wave in a Bose-Hubbard chain. *Phys. Rev. Lett.* **115**, 035302 (2015).
18. DiVincenzo, D. P., Bacon, D., Kempe, J., Burkard, G. & Whaley, K. B. Universal quantum computation with the exchange interaction. *Nature* **408**, 339–342 (2000).
19. Petta, J. R. *et al.* Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).
20. Trotzky, S. *et al.* Time-resolved observation and control of superexchange interactions with ultracold atoms in optical lattices. *Science* **319**, 295–299 (2008).
21. Kotler, S., Akerman, N., Navon, N., Glickman, Y. & Ozeri, R. Measurement of the magnetic interaction between two bound electrons of two separate ions. *Nature* **510**, 376–380 (2014).
22. Daley, A. J., Pichler, H., Schachenmayer, J. & Zoller, P. Measuring entanglement growth in quench dynamics of bosons in an optical lattice. *Phys. Rev. Lett.* **109**, 020505 (2012).
23. Greif, D., Uehlinger, T., Jotzu, G., Tarruell, L. & Esslinger, T. Short-range quantum magnetism of ultracold fermions in an optical lattice. *Science* **340**, 1307–1310 (2013).
24. Murmann, S. *et al.* Two fermions in a double well: exploring a fundamental building block of the Hubbard model. *Phys. Rev. Lett.* **114**, 080402 (2015).
25. Hart, R. A. *et al.* Observation of antiferromagnetic correlations in the Hubbard model with ultracold atoms. *Nature* **519**, 211–214 (2015).
26. Schlosser, N., Reymond, G., Protsenko, I. & Grangier, P. Sub-poissonian loading of single atoms in a microscopic dipole trap. *Nature* **411**, 1024–1027 (2001).
27. Kaufman, A. M., Lester, B. J. & Regal, C. A. Cooling a single atom in an optical tweezer to its quantum ground state. *Phys. Rev. X* **2**, 041014 (2012).
28. Kaufman, A. M. *et al.* Two-particle quantum interference in tunnel-coupled optical tweezers. *Science* **345**, 306–309 (2014).
29. Kielpinski, D. *et al.* A decoherence-free quantum memory using trapped ions. *Science* **291**, 1013–1015 (2001).
30. Wall, M. L., Hazzard, K. R. A. & Rey, A. M. Effective many-body parameters for atoms in nonseparable Gaussian optical potentials. *Phys. Rev. A* **92**, 013610 (2015).
31. Weitenberg, C. *et al.* Single-spin addressing in an atomic Mott insulator. *Nature* **471**, 319–324 (2011).
32. Stewart, G. R. Heavy-fermion systems. *Rev. Mod. Phys.* **56**, 755–787 (1984).

**Acknowledgements** This work was supported by the David and Lucile Packard Foundation and the National Science Foundation under grant number 1125844. C.A.R. acknowledges support from the Clare Boothe Luce Foundation. M.L.W. and A.M.R. acknowledge funding from NSF-PIF, ARO, ARO-DARPA-OLE and AFOSR. M.L.W. and M.F.-F. acknowledge support from the NRC postdoctoral fellowship program.

**Author Contributions** A.M.K. and B.J.L. took the data and performed the data analysis, with guidance from C.A.R. M.F.F., M.L.W. and A.M.R. provided theoretical support. All authors contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.A.R. ([regal@colorado.edu](mailto:regal@colorado.edu)).

## METHODS

**State preparation and experimental protocol.** The experiment begins with atoms loaded stochastically from a magneto-optical trap into a pair of optical tweezers separated by 1.57  $\mu\text{m}$ , each with an optical waist of 710 nm and a 23(1) MHz depth<sup>26</sup>. For all data presented in the figures, we post-select on experiments in which each optical tweezer is loaded with a single atom, on the basis of atom population (0 or 1) measurements in each tweezer at the start of the experiment. We perform optical pumping and 3D Raman sideband cooling as described<sup>27,28</sup>, leaving each atom in the motional ground state of its respective tweezer with 0.90(7) fidelity, as determined via motional sideband spectroscopy, and in the  $|F, m_F\rangle = |2, 2\rangle \equiv |\uparrow\rangle$  spin state. To perform single-atom addressing, we subsequently apply circularly polarized light to the left tweezer to induce a relative shift between the tweezers of the  $|\uparrow\rangle \leftrightarrow |1, 1\rangle \equiv |\downarrow\rangle$  transition<sup>28,31</sup>. For the data in Figs 2b and 3d, we apply a resonant square microwave  $\pi$  pulse to the non-light-shifted atom, yielding a preparation fidelity in the desired two-particle spin configuration of 0.835(10). For the data in Fig. 3c, we improve this fidelity by using temporally shaped Gaussian pulses in order to minimize  $C_{\text{bnd}}$  via the suppression of off-resonant transitions.

After motional and spin preparation, the position of the left optical tweezer is swept adiabatically in 10 ms to the other tweezer, realizing a two-atom spacing (Gaussian centroid separation) of 715 nm (854 nm). The tweezer depths are then reduced to 91(4) kHz in an adiabatic ramp<sup>28</sup>; in this configuration, there is a measured resonant tunnelling of  $J_{\text{eg}}/(2\pi\hbar) = 165(6)$  Hz between the left-tweezer ground state and right-tweezer excited state.

We then apply the first adiabatic passage (AP) through the tunnelling resonance. Starting with the left optical tweezer tuned 2.2 kHz below the tunnelling resonance, the left tweezer depth is swept linearly in 12 ms to symmetrically above the resonance. The end of this ramp marks the nominal  $t = 0$  of the exchange dynamics, and the resulting tweezer depths are static during the exchange time. After the exchange time, the AP is performed exactly in reverse, transferring the motional-excited atom back into its origin tweezer. The full AP procedure in both directions has a measured fidelity of 0.81(4) for the data in Fig. 2b; we observe systematic fluctuations in this fidelity such that its value is 0.69(2) for the data in Fig. 3c.

At this point, if we are directly measuring the exchange dynamics as in Fig. 2, the tweezers are ramped up in depth and separated for spin detection and imaging. If we are performing entanglement detection, instead directly we apply the magnetic-field gradient. After switching on the gradient there is a 3-ms delay for field settling. Then we hold the gradient on for variable time  $t_g$ , after which the gradient is switched off. After another 3-ms delay to let the fields settle, we apply the microwave  $\pi/2$  pulse for parity read-out, after which the tweezers are increased in depth and separated for spin detection and imaging.

The observed spin-exchange contrast can be influenced by a number of factors, including the single-particle ground-state fraction, single-particle loss, and the AP fidelities. Spin preparation also affects the contrast, but as discussed in the text the data presented in Fig. 2b are post-selected on the desired anti-aligned spin configuration (for the parity measurements, post-selection is not possible since we study all final spin configurations). On the basis of our measurements of experimental systematics—the AP fidelity of 0.81(4), the single-particle ground-state fraction of 0.90(7), and the single-particle survival probability of 0.963(7)—and combined with the spin-preparation post-selection, we would expect a spin-exchange contrast of 0.60(10), which exceeds the measured value of 0.29(2) in Fig. 2b. In the absence of the spin-preparation post-selection, we observe a spin-exchange contrast that is reduced by the spin-preparation fidelity of 0.835(10), as expected.

For the characterization of the spin populations, in a distinct experiment we measure the spin-down probability of the atom in each tweezer by applying resonant push-out light directly before when the first AP of the experiment would occur. Accounting for the separately measured single-atom survival probability of 0.963(7), we measure  $\{P^{\uparrow\uparrow}, P^{\downarrow\downarrow}\} = \{0.071(14), 0.016(5)\}$ , leading to  $P^{\uparrow\downarrow} + P^{\downarrow\uparrow} = 1 - (P^{\uparrow\uparrow} + P^{\downarrow\downarrow})$ . These numbers are used to calculate  $C_{\text{bnd}}$  and  $F$ . The measured parity data are not corrected for loss of single atoms due to background collisions, which degrades the entanglement generation.

**Modelling of interactions and tunnelling during state preparation.** For two atoms of mass  $m$  in different motional wave functions  $\psi_i(\mathbf{r}) = \langle \mathbf{r} | e \rangle$  and  $\psi_j(\mathbf{r}) = \langle \mathbf{r} | g \rangle$ , the exchange interaction energy  $J_{\text{ex}} = 2U_{\text{eg}}$ , where

$U_{\text{eg}} = \frac{4\pi\hbar^2 a_s}{m} \int |\psi_e(\mathbf{r})|^2 |\psi_g(\mathbf{r})|^2 d^3\mathbf{r}$  is the contact interaction energy proportional to  $a_s$ , the atomic  $s$ -wave scattering length. To compute this integral we use efficient numerical techniques to determine the spectrum and eigenstates of the 3D optical tweezer potential<sup>30</sup> using parameters from an independent experimental characterization<sup>28</sup>. This calculation, combined with experimental uncertainties, gives rise to the blue theoretical swath in Fig. 2.

To theoretically model the AP procedure, we consider a situation in which two particles with opposite spins are initially localized in the ground motional states of the left and right wells, denoted  $L_g$  and  $R_g$ , respectively. During the AP, the bias between wells  $\Delta$  is tuned so that  $L_g$  is near resonance with an excited level of the right well, denoted  $R_e$ . In what follows, we will measure the bias  $\Delta$  with respect to this resonance position, that is, the resonance occurs at  $\Delta = 0$ . Further, we consider that the bias range is such that tunnel couplings to all other motional states (for example,  $L_g \rightarrow R_e$  tunnelling) are negligible, and so we can restrict ourselves to the set of motional states spanned by  $L_g, R_g$  and  $R_e$  and the single  $L_g \rightarrow R_e$  tunnelling resonance at  $\Delta = 0$ . Because of the mixing of spin components due to the spin-exchange interaction, it is most convenient to use the singlet-triplet basis  $\{|S; L_g, R_g\rangle, |S; R_e, R_g\rangle, |T; L_g, R_g\rangle, |T; R_e, R_g\rangle\}$ , where  $S$  and  $T$  denote singlet and triplet states and the last two labels are the motional states of the two particles. In this basis, the Hamiltonian is

$$\hat{H} = \begin{pmatrix} \Delta & -J_{\text{eg}} & 0 & 0 \\ -J_{\text{eg}} & 0 & 0 & 0 \\ 0 & 0 & \Delta & -J_{\text{eg}} \\ 0 & 0 & -J_{\text{eg}} & 2U_{\text{eg}} \end{pmatrix} \quad (2)$$

where  $J_{\text{eg}}$  is the tunnelling amplitude for the process  $L_g \rightarrow R_e$ , and the initial state at large negative  $\Delta$ ,  $|L_g\uparrow, R_g\downarrow\rangle$ , is an equal weight superposition of  $|S; L_g, R_g\rangle$  and  $|T; L_g, R_g\rangle$ . The AP process is described by separate tunnelling avoided crossings in the singlet and triplet channels, as shown in Extended Data Fig. 1. Provided that the ramping procedure is adiabatic with respect to these avoided crossings, it will transfer the initial state into an equal weight superposition of  $|S; R_e, R_g\rangle$  and  $|T; R_e, R_g\rangle$  at large positive  $\Delta$ . This pair of eigenstates, which correspond to the two particles occupying the same well, have an asymptotic energy splitting of  $2U_{\text{eg}}$  as  $|\Delta| \gg J_{\text{eg}}$ . Note that the position of the bias resonance for the triplet channel is shifted by  $2U_{\text{eg}}$  with respect to the resonance in the singlet channel. The energy offset does not affect the degree of adiabaticity of the ramping procedure and just gives rise to a phase shift between the singlet and triplet components.

**Entanglement verification based on parity oscillations.** Here we derive a criterion for verifying entanglement generated by spin-exchange interactions in a two-atom system. Our strategy will be to assume an unentangled (separable) density matrix, and from this assumption establish a constraint on experimentally measurable quantities: the parity oscillation contrast (Fig. 3c) and the populations of different spin states. Experimental violation of this constraint thus verifies entanglement.

In the experiment, spin exchange occurs between two atoms occupying a single optical tweezer, and then those atoms are separated into two tweezers. Because the APs in the experiment are imperfect, the atoms may sometimes end up in the same tweezer after the attempted separation. For clarity of presentation, we first consider the idealized case of perfect AP fidelity; hence  $\rho$  in what follows describes states immediately after the second AP in which one atom occupies each tweezer. At the end of this section we carefully consider the effects of AP failure, and show that they do not affect our claims of entanglement.

Because the measured spin-coherence time in the experiment is much less than the time between when we prepare the initial spin state and when we complete the APs,  $\rho$  cannot have any coherences between states with different total  $S^z = S_L^z + S_R^z$ . Working in a basis that diagonalizes both  $\hat{S}_L^z$  and  $\hat{S}_R^z$ ,  $\{|\uparrow\uparrow\rangle_L, |\uparrow\downarrow\rangle_L, |\downarrow\uparrow\rangle_L, |\downarrow\downarrow\rangle_L, |\uparrow\uparrow\rangle_R, |\uparrow\downarrow\rangle_R, |\downarrow\uparrow\rangle_R, |\downarrow\downarrow\rangle_R\}$ , the most general density matrix satisfying this condition can be written

$$\rho = \begin{pmatrix} P(\uparrow_L, \uparrow_R) & 0 & 0 & 0 \\ 0 & P(\uparrow_L, \downarrow_R) & \epsilon & 0 \\ 0 & \epsilon^* & P(\downarrow_L, \uparrow_R) & 0 \\ 0 & 0 & 0 & P(\downarrow_L, \downarrow_R) \end{pmatrix} \quad (3)$$

The populations  $P(\uparrow_L, \uparrow_R)$  and  $P(\downarrow_L, \downarrow_R)$  are, respectively, the total probabilities of having both atoms in the  $|\uparrow\rangle$  state or both atoms in the  $|\downarrow\rangle$  state, and are non-zero owing to imperfect initial spin preparation. Because these probabilities are conserved by the APs and the spin exchange, their measured values before the AP, referred to as  $P^{\uparrow\uparrow}$  and  $P^{\downarrow\downarrow}$  in the main text, can safely be used in equation (3):  $P(\uparrow_L, \uparrow_R) = P^{\uparrow\uparrow}$  and  $P(\downarrow_L, \downarrow_R) = P^{\downarrow\downarrow}$ . The parity is measured after first applying a magnetic-field gradient for a variable time  $t_g$ , and then applying a  $\pi/2$  microwave pulse, which transforms  $\rho \rightarrow \tilde{\rho}(t_g)$ . After some algebra, the parity of  $\tilde{\rho}(t_g)$  can be written  $\Pi(t_g) = 2\text{Re}(e^{-i\delta t_g} \epsilon)$ , which oscillates as a function of  $t_g$  with a contrast of  $C = 4|\epsilon|$ .

Our goal is to derive a constraint on  $C$  in terms of the measured quantities  $P^{\uparrow\uparrow}$ ,  $P^{\downarrow\downarrow}$ , under the assumption that  $\rho$  is separable. If  $\rho$  were a product state  $\rho_L \otimes \rho_R$ , where

$$\rho_{L(R)} = \begin{pmatrix} \rho_{L(R)}^{\uparrow\uparrow} & \rho_{L(R)}^{\uparrow\downarrow} \\ \rho_{L(R)}^{\downarrow\uparrow} & \rho_{L(R)}^{\downarrow\downarrow} \end{pmatrix} \quad (4)$$

then

$$|\varepsilon| = |\rho_L^{\uparrow\downarrow} \rho_R^{\downarrow\uparrow}| \leq (\rho_L^{\uparrow\uparrow} \rho_L^{\downarrow\downarrow})^{1/2} \times (\rho_R^{\uparrow\uparrow} \rho_R^{\downarrow\downarrow})^{1/2} = (P(\uparrow_L, \uparrow_R) P(\downarrow_L, \downarrow_R))^{1/2} = (P^{\uparrow\uparrow} P^{\downarrow\downarrow})^{1/2} \quad (5)$$

By definition, a separable state can be written as a classical mixture of product states,  $\rho = \sum_j \lambda_j \rho_L^j \otimes \rho_R^j$ , in which case

$$|\varepsilon| \leq \sum_j \lambda_j |\varepsilon_j| \leq \sum_j \lambda_j (P_j^{\uparrow\uparrow} P_j^{\downarrow\downarrow})^{1/2} \leq \left( \sum_j \lambda_j P_j^{\uparrow\uparrow} \right)^{1/2} \left( \sum_j \lambda_j P_j^{\downarrow\downarrow} \right)^{1/2} = (P^{\uparrow\uparrow} P^{\downarrow\downarrow})^{1/2} \quad (6)$$

Here, the second inequality is the constraint on  $|\varepsilon|$  derived for a product state in equation (5) applied to each state in the classical mixture, and the first and third are triangle inequalities. Using  $C = 4|\varepsilon|$ , we therefore have guaranteed entanglement whenever

$$C > 4(P^{\uparrow\uparrow} P^{\downarrow\downarrow})^{1/2} \quad (7)$$

We can also verify entanglement by using an entanglement witness associated with the fidelity of  $\rho$  in the maximally entangled state  $|\Psi_+\rangle = (|\uparrow\rangle_L |\downarrow\rangle_R + |\downarrow\rangle_L |\uparrow\rangle_R) / \sqrt{2}$ ,  $F = \langle \Psi_+ | \rho | \Psi_+ \rangle$ . As shown in ref. 1,  $\rho$  is entangled if  $F > 1/2$ . In terms of equation (3), the fidelity can be written

$$F = \frac{1}{2} [P(\uparrow_L, \downarrow_R) + P(\downarrow_L, \uparrow_R)] + \frac{C}{4} = \frac{1}{2} + \frac{C}{4} - \frac{1}{2} (P^{\uparrow\uparrow} + P^{\downarrow\downarrow}) \quad (8)$$

Therefore  $F > 1/2$  is equivalent to  $C > 2(P^{\uparrow\uparrow} + P^{\downarrow\downarrow})$ , which agrees with equation (7) when  $P^{\uparrow\uparrow} = P^{\downarrow\downarrow}$ . Equation (7) can also be derived by applying the Peres–Horodecki criterion to  $\rho$ , and thus it is actually both sufficient and necessary for entanglement<sup>33,34</sup>. Thus, in contrast with the fidelity-based entanglement witness  $C > 2(P^{\uparrow\uparrow} + P^{\downarrow\downarrow})$ , the right-hand-side of equation (7) is as small as possible for any  $P^{\uparrow\uparrow}$  and  $P^{\downarrow\downarrow}$ , resulting in the greatest possible confidence in entanglement for a particular measured contrast. We also note that the extent to which equation (7) is satisfied,  $\frac{1}{2}(C - 4(P^{\uparrow\uparrow} P^{\downarrow\downarrow})^{1/2}) > 0$ , is a direct measurement of the concurrence in the density matrix  $\rho$  (ref. 16).

**Imperfect adiabatic passages.** As mentioned above, the APs in the experiment are not perfect. For our purposes, we define success as any situation in which the

atoms end up in different tweezers, which occurs either if both of the individual APs succeed or if they both fail. Conversely, we define failure as any situation in which both atoms end up in the same tweezer, which happens if one of the APs is successful while the other is not. Although the states resulting from failure have not been considered in deriving equation (7), they do not contribute to the parity oscillation contrast, since two atoms in the same tweezer are not sensitive to a magnetic-field gradient. Therefore, we intuitively expect that imperfect AP can only lower the measured contrast, such that equation (7) still implies entanglement.

This intuition can be formalized by introducing a projector,  $\hat{K}$ , onto the states with one atom in each tweezer. Defining  $f$  as the success probability, we can then form projections of the true experimental density matrix, denoted by  $\rho_{\text{exp}}$ , into the subspaces defined by success or failure

$$\hat{K} \rho_{\text{exp}} \hat{K} \equiv f \rho_{\text{succ}} \quad \text{and} \quad (1 - \hat{K}) \rho_{\text{exp}} (1 - \hat{K}) \equiv (1 - f) \rho_{\text{fail}} \quad (9)$$

Note that, by the choice of pre-factors, both  $\rho_{\text{succ}}$  and  $\rho_{\text{fail}}$  are properly normalized density matrices. Importantly,  $\rho_{\text{succ}}$  has precisely the form given in equation (3); we can therefore directly apply the arguments above, deducing that  $\rho_{\text{succ}}$  is entangled whenever  $|\varepsilon| > \sqrt{P^{\uparrow\uparrow} P^{\downarrow\downarrow}}$ . The only difference from before is in the relationship between  $|\varepsilon|$  and the measured parity oscillation contrast. The application of a magnetic-field gradient and the subsequent  $\pi/2$  pulse does not mix the two subspaces partitioned by  $\hat{K}$ , and therefore the measured parity can be written  $\Pi(t_g) = f \Pi_{\text{succ}}(t_g) + (1 - f) \Pi_{\text{fail}}$ . Importantly,  $\Pi_{\text{fail}}$  is independent of  $t_g$  because two atoms in the same tweezer are not sensitive to a magnetic-field gradient, and therefore the measured parity oscillation contrast is given simply by  $C = f 4|\varepsilon|$ .

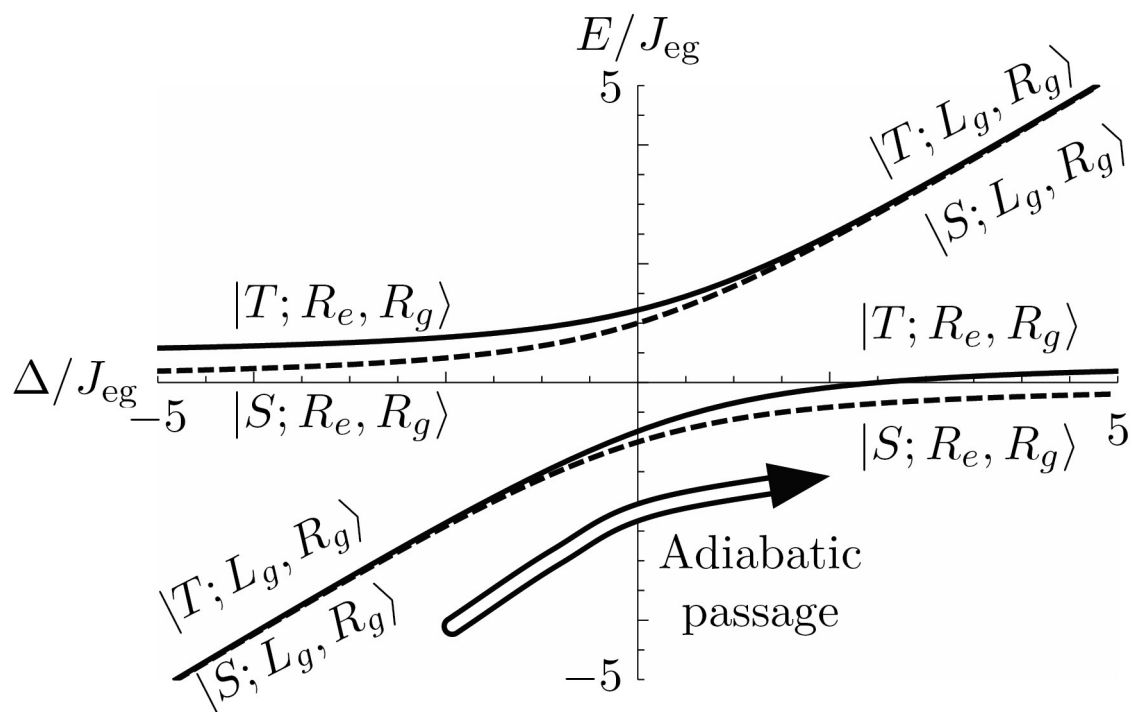
Thus the effect of imperfect APs is that, in terms of the measured contrast, the condition  $|\varepsilon| > \sqrt{P^{\uparrow\uparrow} P^{\downarrow\downarrow}}$  now reads  $C/f > 4\sqrt{P^{\uparrow\uparrow} P^{\downarrow\downarrow}}$  (compare with equation (7)).

The criterion used in the main text,  $C > 4\sqrt{P^{\uparrow\uparrow} P^{\downarrow\downarrow}}$ , is therefore a conservative way to verify entanglement in  $\rho_{\text{succ}}$ , since  $C/f$  is strictly larger than  $C$ .

The fidelity can now be written  $F = \langle \Psi_+ | \rho_{\text{exp}} | \Psi_+ \rangle = f \langle \Psi_+ | \rho_{\text{succ}} | \Psi_+ \rangle \equiv f F_{\text{succ}}$ , where  $F_{\text{succ}}$  is the fidelity of the (projected and normalized) density matrix  $\rho_{\text{succ}}$ , and can be equivalently written as  $F_{\text{succ}} = \frac{1}{2} + |\varepsilon| - \frac{1}{2} (P^{\uparrow\uparrow} + P^{\downarrow\downarrow})$ . Similar to before, entanglement of  $\rho_{\text{succ}}$  is now guaranteed by  $F_{\text{succ}} > 1/2$ , or equivalently  $4|\varepsilon| > 2(P^{\uparrow\uparrow} + P^{\downarrow\downarrow})$ . Using  $4|\varepsilon| = C/f$ , we can extract  $F_{\text{succ}}$  from the measured contrast ( $C = 0.49(4)$ ) and measured success probability ( $f = 0.69(2)$ ), obtaining  $F_{\text{succ}} = 0.634(17)$ . Note that the actual fidelity  $F$  is reduced from this value by the success probability  $f$ , and is in fact below  $1/2$ . However, this is not inconsistent with entanglement in  $\rho_{\text{exp}}$ , as the fidelity-based witness for  $\rho_{\text{exp}}$ , written in terms of the actual fidelity, is  $F > f/2$ .

33. Peres, A. Separability criterion for density matrices. *Phys. Rev. Lett.* **77**, 1413–1415 (1996).
34. Horodecki, M., Horodecki, P. & Horodecki, R. Separability of mixed states: necessary and sufficient conditions. Preprint at <http://arXiv.org/abs/quant-ph/9605038> (1996).





**Extended Data Figure 1 |** Adiabatic energy eigenstates  $E$  as a function of the double-well bias  $\Delta$  in units of the ground-excited tunnelling  $J_{eg}$ . At large positive bias, the triplet and singlet eigenstates corresponding to two

particles in the same well are split by  $J_{ex}$ . The dashed and solid lines denote the energies of the states that asymptotically connect to the states labelled in the figure through the AP process. See Methods for details.

# Large anomalous Hall effect in a non-collinear antiferromagnet at room temperature

Satoru Nakatsuji<sup>1,2</sup>, Naoki Kiyohara<sup>1</sup> & Tomoya Higo<sup>1</sup>

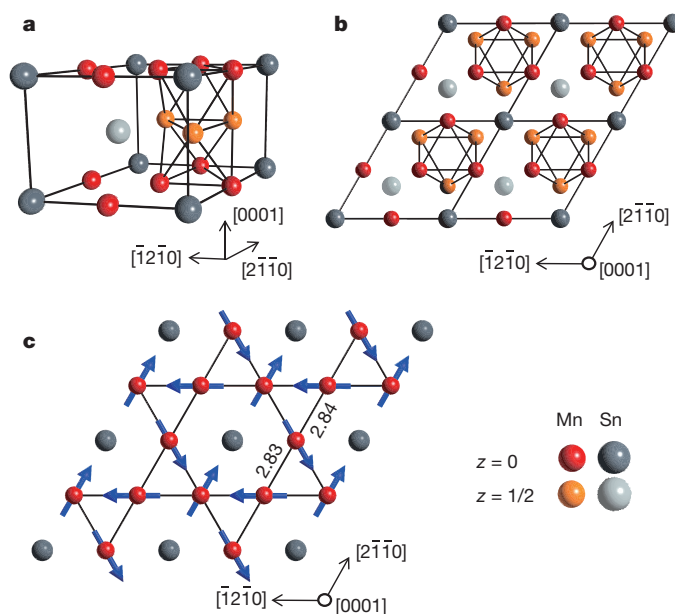
In ferromagnetic conductors, an electric current may induce a transverse voltage drop in zero applied magnetic field: this anomalous Hall effect<sup>1</sup> is observed to be proportional to magnetization, and thus is not usually seen in antiferromagnets in zero field<sup>2</sup>. Recent developments in theory and experiment have provided a framework for understanding the anomalous Hall effect using Berry-phase concepts<sup>3</sup>, and this perspective has led to predictions that, under certain conditions, a large anomalous Hall effect may appear in spin liquids and antiferromagnets without net spin magnetization<sup>4–8</sup>. Although such a spontaneous Hall effect has now been observed in a spin liquid state<sup>9</sup>, a zero-field anomalous Hall effect has hitherto not been reported for antiferromagnets. Here we report empirical evidence for a large anomalous Hall effect in an antiferromagnet that has vanishingly small magnetization. In particular, we find that  $\text{Mn}_3\text{Sn}$ , an antiferromagnet that has a non-collinear 120-degree spin order<sup>10,11</sup>, exhibits a large anomalous Hall conductivity of around 20 per ohm per centimetre at room temperature and more than 100 per ohm per centimetre at low temperatures, reaching the same order of magnitude as in ferromagnetic metals<sup>3</sup>. Notably, the chiral antiferromagnetic state has a very weak and soft ferromagnetic moment of about 0.002 Bohr magnetons per Mn atom (refs 10, 12), allowing us to switch the sign of the Hall effect with a small magnetic field of around a few hundred oersted. This soft response of the large anomalous Hall effect could be useful for various applications including spintronics—for example, to develop a memory device that produces almost no perturbing stray fields.

$\text{Mn}_3\text{Sn}$  is a hexagonal antiferromagnet (AFM) that exhibits non-collinear ordering of Mn magnetic moments at the Néel temperature of  $T_N \approx 420$  K (refs 10, 11, 13). The system has a hexagonal  $\text{Ni}_3\text{Sn}$ -type structure with space group  $P6_3/mmc$  (Fig. 1a). The structure is stable only in the presence of excess Mn, which randomly occupies the Sn site<sup>13</sup>. The basal plane projection of the Mn sublattice can be viewed as a triangular lattice arrangement of a twisted triangular tube made of face-sharing octahedra (Fig. 1a, b). Each  $a$ - $b$  plane consists of a slightly distorted kagome lattice of Mn moments each of  $\sim 3 \mu_B$  (where  $\mu_B$  is the Bohr magneton), and the associated geometrical frustration manifests itself as an inverse triangular spin structure that carries a very small net ferromagnetic moment of  $\sim 0.002 \mu_B$  per Mn atom (Fig. 1c)<sup>10,11</sup>. All Mn moments lie in the  $a$ - $b$  plane and form a chiral spin texture with an opposite vector chirality to the usual 120° structure (Extended Data Fig. 1). This inverse triangular configuration has an orthorhombic symmetry, and only one of the three moments in each Mn triangle is parallel to the local easy-axis<sup>10–12</sup> (Fig. 1c). Thus, the canting of the other two spins towards the local easy-axis is considered to be the origin of the weak ferromagnetic moment<sup>10–12</sup>.

It is known that as-grown crystals retain the inverse triangular spin state over a wide temperature ( $T$ ) range between  $T_N$  and  $\sim 50$  K (ref. 14). At low temperatures, a cluster glass phase appears with a large  $c$ -axis ferromagnetic component due to spin canting towards the  $c$  axis<sup>11,15,16</sup>. In this work, we used as-grown single crystals that have the

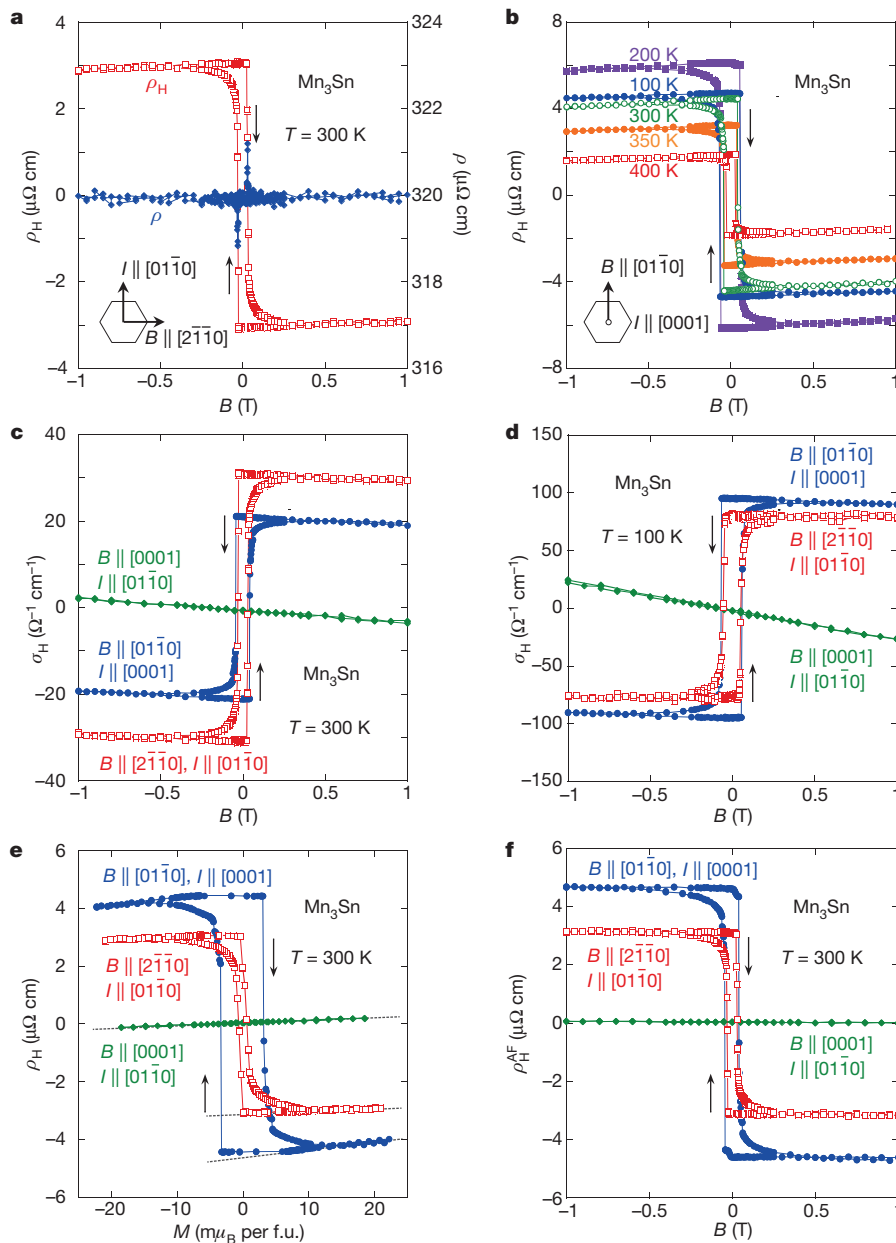
composition  $\text{Mn}_{3.02}\text{Sn}_{0.98}$  and confirmed no transition except the one at 50 K (Methods). As the detailed spin structure is unknown for the low temperature phase, here we focus on the phase stable above 50 K, and use ' $\text{Mn}_3\text{Sn}$ ' to refer to our crystals for clarity.

We first show our main experimental evidence for the large anomalous Hall effect (AHE) at room temperature. Figure 2a presents the field dependence of the Hall resistivity,  $\rho_H(B)$ , obtained at 300 K for the field along  $[2\bar{1}10]$  ( $a$  axis).  $\rho_H(B)$  exhibits a clear hysteresis loop with a sizable jump of  $|\Delta\rho_H| \approx 6 \mu\Omega \text{ cm}$ . This is strikingly large for an AFM, and is larger than those found in elemental transition metal ferromagnets (FMs) such as Fe, Co and Ni (refs 2, 3, 17). Notably, the sign change occurs at a small field of  $\sim 300$  Oe. Furthermore, the hysteresis remains sharp and narrow in all the temperature range between 100 K and 400 K (Fig. 2b). In this temperature region, a large anomaly as a function of field has been seen only in the Hall resistivity. The longitudinal resistivity  $\rho(B)$  remains constant except for spikes at the critical fields where the Hall resistivity jumps (Fig. 2a).



**Figure 1 | Crystal and magnetic structures of  $\text{Mn}_3\text{Sn}$ .** **a**, The crystallographic unit cell of  $\text{Mn}_3\text{Sn}$ . Although there is only one crystallographic site for both Mn and Sn, different colours are used to distinguish those in the  $z=0$  plane and in the  $z=1/2$  plane. In addition to the unit cell frame, Mn atoms are connected by lines to illustrate that the face-sharing octahedra of Mn atoms form a twisted triangular tube along the  $c$  axis. **b**, Top view along the  $c$  axis of the neighbouring four unit cells in the  $a$ - $b$  plane. **c**, An individual  $a$ - $b$  plane of  $\text{Mn}_3\text{Sn}$ . All distances are in Å. Mn moments (arrows) form an inverse triangular spin structure<sup>10–12</sup>. Each Mn moment has the local easy-axis parallel to the in-plane direction towards its nearest-neighbour Sn sites. Here,  $[2\bar{1}10]$ ,  $[\bar{1}2\bar{1}0]$  and  $[0001]$  are the  $a$ ,  $b$  and  $c$  axes, respectively

<sup>1</sup>Institute for Solid State Physics, University of Tokyo, Kashiwa 277-8581, Japan. <sup>2</sup>PRESTO, Japan Science and Technology Agency (JST), 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan.



**Figure 2 | Magnetic field dependence of the AHE in  $\text{Mn}_3\text{Sn}$ .** **a**, Field dependence of the Hall resistivity  $\rho_H$  (left axis) and the longitudinal resistivity  $\rho$  (right axis) at 300 K in the magnetic field  $B \parallel [2\bar{1}10]$  with the electric current  $I \parallel [01\bar{1}0]$ . **b**, Field dependence of the Hall resistivity  $\rho_H$  at various temperatures in  $B \parallel [01\bar{1}0]$  with  $I \parallel [0001]$ . **c, d**, The Hall conductivity  $\sigma_H$  versus  $B$  measured in  $B \parallel [2\bar{1}10]$ ,  $[01\bar{1}0]$  and  $[0001]$  obtained at 300 K (**c**) and 100 K (**d**). **e**, Magnetization dependence of  $\rho_H$  at 300 K. **f**, Field dependence of  $\rho_H^{\text{AF}} = \rho_H - R_0 B - R_s \mu_0 M$  at 300 K. The arrows in the hexagon at lower left in **a** and **b** indicate the field and current directions in the hexagonal lattice of  $\text{Mn}_3\text{Sn}$ .

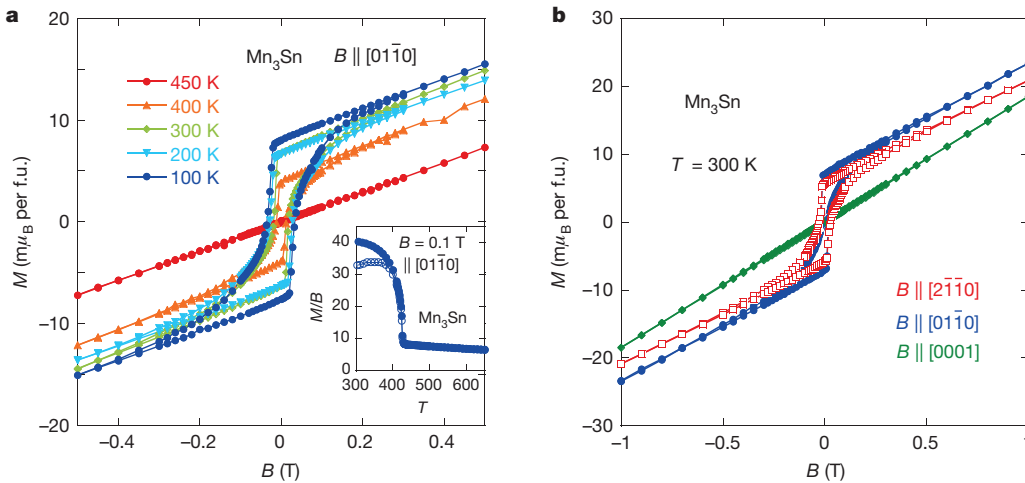
Correspondingly, the Hall conductivity,  $\sigma_H = -\rho_H/\rho^2$ , for in-plane fields along both  $[2\bar{1}10]$  and  $[01\bar{1}0]$  shows a large jump and narrow hysteresis (Fig. 2c, d). For instance, with  $B \parallel [01\bar{1}0]$ ,  $\sigma_H$  has large values near zero field,  $\sim 20 \Omega^{-1} \text{cm}^{-1}$  at 300 K and nearly  $100 \Omega^{-1} \text{cm}^{-1}$  at 100 K. This is again quite large for an AFM and comparable to those values found in ferromagnetic metals<sup>3,17</sup>. On the other hand, the Hall conductivity for  $B \parallel [0001]$  (*c* axis) shows no hysteresis but only a linear field dependence.

The magnetization curve  $M(B)$  shows anisotropic hysteresis similar to that found for the Hall effect. For example,  $M$  versus  $B \parallel [01\bar{1}0]$  at temperatures between 100 K and 400 K shows a clear hysteresis, indicating that a weak ferromagnetic moment ( $4\text{--}7 \text{ m}\mu_B$  per formula unit (f.u.)) changes its direction with coercivity of only a few hundred oersted (Fig. 3a). Whereas the in-plane  $M$  is almost isotropic and has a narrow hysteresis, the magnetization shows only a linear dependence on  $B$  for  $B \parallel [0001]$  at all the temperatures measured between 100 K and 450 K (Fig. 3b). The similar anisotropic and hysteretic behaviours found in both  $\rho_H(B)$  and  $M(B)$  indicate that the existence of the small and soft ferromagnetic component allows us to switch the sign of the Hall effect. Indeed, previous neutron diffraction measurements and theoretical analyses clarified that the inverse triangular spin structure has no

in-plane anisotropy energy up to the fourth-order term<sup>10,12</sup>, which is consistent with the observed small coercivity. This further indicates that by rotating the net ferromagnetic moment, one may switch the staggered moment direction of the triangular spin structure<sup>10,12</sup>. This switch should be the origin of the sign change of the Hall effect, as we discuss below. On heating, this ferromagnetic component vanishes at the Néel temperature of 430 K, above which the hysteresis disappears in both the  $T$  and  $B$  dependence of the magnetization (Fig. 3a and its inset).

To reveal the temperature evolution of the spontaneous component of the AHE, both the zero-field Hall resistivity  $\rho_H(B=0)$  and the zero-field longitudinal resistivity  $\rho(B=0)$  were measured after cooling samples in a magnetic field of  $B_{\text{FC}} = 7 \text{ T}$  from 400 K down to 5 K and subsequently setting  $B$  to 0 at 5 K (Methods). Figure 4a shows the temperature dependence of the zero-field Hall conductivity  $\sigma_H(B=0) = -\rho_H(B=0)/\rho^2(B=0)$  obtained after the above field-cooling (FC) procedure using three different configurations of the magnetic field ( $B_{\text{FC}}$ ) and electric current ( $I$ ) directions. Here,  $\sigma_{zx}$  stands for the Hall conductivity obtained after the FC procedure in  $B_{\text{FC}} \parallel [01\bar{1}0]$  with  $I \parallel [0001]$ , and  $\sigma_{yz}$  for  $B_{\text{FC}} \parallel [2\bar{1}10]$  and  $I \parallel [01\bar{1}0]$ . Both show large values at low temperatures, and in particular,  $|\sigma_{zx}|$  exceeds  $100 \Omega^{-1} \text{cm}^{-1}$  at  $T < 80 \text{ K}$ . Both  $|\sigma_{zx}|$  and  $|\sigma_{yz}|$  decrease on heating but still retain





**Figure 3 | In-plane weak ferromagnetism in Mn<sub>3</sub>Sn.** **a**, Field dependence of the magnetization  $M$  in the magnetic field  $B \parallel [01\bar{1}0]$  at various temperatures. Inset shows the temperature (in K) dependence of the susceptibility  $M/B$  (in milli-electromagnetic units per mol per f.u.) obtained above 300 K in the field of 0.1 T  $\parallel [01\bar{1}0]$ . Filled and open symbols indicate the results obtained in the field-cooling and zero-field-cooling procedures, respectively (see Methods). **b**, Anisotropy in the magnetization curve obtained at 300 K in  $B \parallel [2\bar{1}10]$ ,  $[01\bar{1}0]$  and  $[0001]$ .

values of  $\sim 10 \Omega^{-1} \text{ cm}^{-1}$  at 400 K, which is the highest temperature of our measurements. On the other hand,  $\sigma_{xy}$  obtained after the FC procedure in  $B_{FC} \parallel [0001]$  with  $I \parallel [01\bar{1}0]$  is zero within our experimental accuracy at  $T > 50$  K. In the low temperature phase below 50 K,  $|\sigma_{xy}|$  increases on cooling and reaches  $140 \Omega^{-1} \text{ cm}^{-1}$  at 5 K, the lowest temperature of our measurements. In the three FC procedures described above, the temperature dependence of the longitudinal resistivity  $\rho(B=0)$  was also concomitantly obtained (Fig. 4a, inset). Both in-plane and out-of-plane components show saturation at  $T > 300$  K, indicating the presence of strong inelastic scattering at high temperatures.

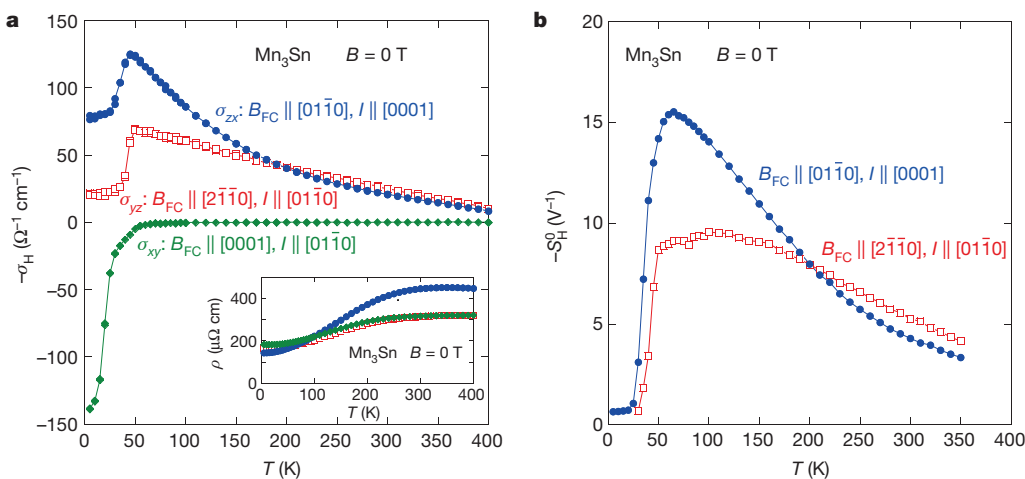
Conventionally, the Hall resistivity is described as  $\rho_H = R_0 B + R_s \mu_0 M$ . Here,  $R_0$  and  $R_s$  are the ordinary and anomalous Hall coefficients, and  $\mu_0$  is the permeability. To further examine the field and magnetization dependence of the AHE, we estimated the ordinary Hall contribution  $R_0 B$  by using the temperature dependent  $\rho_H$  and  $M/B$  for  $B \parallel c$  (Extended Data Fig. 2, Methods). The obtained  $R_0 = 3.0 \times 10^{-4} \text{ cm}^3 \text{ C}^{-1}$  indicates that  $R_0 B$  is negligibly small compared to the observed  $\rho_H$ . Plotting  $\rho_H$  versus  $M$  in Fig. 2e, we note that  $\rho_H$  for  $B \parallel c$  has a normal  $M$ -linear AHE. Likewise,  $\rho_H$  for  $B \parallel a-b$  also shows an  $M$ -linear AHE in field,  $\Delta \rho_H = R_s \mu_0 M$  (broken lines). Clearly, however, the large hysteresis with a sharp sign change in  $\rho_H$  cannot be described by the simple linear term, indicating that there is another dominant contribution to the AHE. If we label this additional term as  $\rho_H^{\text{AF}}$ , the Hall resistivity in Mn<sub>3</sub>Sn can be described by

$$\rho_H = R_0 B + R_s \mu_0 M + \rho_H^{\text{AF}} \quad (1)$$

By subtracting  $R_0 B$  and  $R_s \mu_0 M$  from  $\rho_H$ , we find that  $\rho_H^{\text{AF}}$  is nearly independent of  $B$  or  $M$ , unlike what is found in FMs (Fig. 2f, Extended Data Fig. 3). With the reversal of a small applied field,  $\rho_H^{\text{AF}}$  changes sign, corresponding to the rotation of the staggered moments of the non-collinear spin structure<sup>10,12</sup>. Thus, the large AHE,  $\rho_H^{\text{AF}}$ , must have a distinct origin driven by the antiferromagnetic order.

In a magnetic conductor with relatively high resistivity, the AHE is dominated by contributions  $\propto \rho^2$ . Thus, it is useful to compare  $S_H = \mu_0 R_s / \rho^2$  for Mn<sub>3</sub>Sn with those for various magnets (Extended Data Table 1, Methods)<sup>3</sup>. Normally for FMs such as Fe, Ni and MnSi,  $S_H$  is known to be field-independent, and takes values of the order of  $0.01\text{--}0.1 \text{ V}^{-1}$  (refs 3, 18, 19). Indeed, the field-induced  $M$ -linear contribution of the AHE has a field-independent  $S_H$ , which has the positive sign and the same order of magnitude as in FMs. On the other hand, one can also define  $S_H$  for the spontaneous component at zero field as  $S_H^0 = \rho_H(B=0) / [\rho^2(B=0)M(B=0)] = \rho_H^{\text{AF}}(B=0) / [\rho^2(B=0)M(B=0)] + S_H$ . We find significantly large  $|S_H^0| \gg |S_H|$ , reaching  $14 \text{ V}^{-1}$  at 100 K and with a different sign from  $S_H$  (Fig. 4b, Extended Data Table 1). This indicates that  $\rho_H^{\text{AF}}$ , which is the dominant part of the spontaneous component, has a different origin from the conventional AHE (Methods).

A large AHE in a non-collinear AFM was first theoretically predicted for Mn<sub>3</sub>Ir, which has a stacked kagome lattice of Mn atoms, similarly to Mn<sub>3</sub>Sn (ref. 7). Chen *et al.*<sup>7</sup> considered that an AHE may be induced by breaking a symmetry of a single layer kagome lattice that has a triangular



**Figure 4 | Temperature evolution of the zero-field component of the AHE.** Open square, filled circle and filled diamond symbols represent the results obtained after field-cooling (FC) procedures under magnetic field  $B_{FC}$  along the  $[2\bar{1}10]$ ,  $[01\bar{1}0]$  and  $[0001]$  directions, respectively. The directions of the electric current  $I$  are also specified in each main panel. For details, see Methods. **a**, Temperature dependence of the anomalous Hall conductivity,  $\sigma_H(B=0)$ . Inset, temperature dependence of the zero-field resistivity concomitantly obtained after the same FC procedures. **b**, Temperature dependence of  $S_H^0 = -\sigma_H(B=0) / M(B=0)$  obtained after the FC procedures. Here  $M(B=0)$  is the remanent magnetization.

magnetic order, and confirmed the large AHE by numerical calculations. Similar symmetry arguments apply to  $\text{Mn}_3\text{Sn}$ . In this case, the inverse triangular magnetic order breaks the in-plane hexagonal symmetry of the lattice, and thus may induce an AHE in the  $a$ - $b$  plane. Indeed Kübler *et al.*<sup>8</sup> have theoretically found a large AHE in  $\text{Mn}_3\text{Sn}$ , calculating the anomalous Hall conductivity through the Brillouin zone integration of the Berry curvature<sup>3,20</sup>. Interestingly, they found a significant enhancement of the Berry curvature, particularly around band crossing points called Weyl points<sup>21,22</sup> near the Fermi energy. Experimental confirmation of the existence of the Weyl points (for example by ARPES measurements) is awaited. On the other hand, the observed  $\sigma_{xy} = 0$  at  $T > 50$  K is consistent with the in-plane coplanar spin structure and with the absence of a topological Hall effect due to spin chirality<sup>4–6,8</sup>. The large enhancement in  $|\sigma_{xy}|$  at  $T < 50$  K, however, may contain the topological Hall contribution as spins cant towards the  $c$  axis in the low- $T$  phase. It would be interesting to verify this possibility in future studies.

Various applications are conceivable for the observed large AHE and its soft response to applied field. One possibility is for non-volatile memory. To date, FMs have been used as the main active materials for memory devices<sup>23</sup>. However, AFMs have recently attracted attention, because the small magnitude of their stray fields provides stability against magnetic field perturbations, opening new avenues to achieving high data retention and high-density memory integration<sup>24–29</sup>. In addition, AFMs have much faster spin dynamics than FMs, which may lead to ultrafast data processing<sup>26,29</sup>.

To develop useful magnetic devices, one needs to find detectable macroscopic effects that are induced by the rotation of the order parameter. Such effects commonly used for FMs are often unavailable in AFMs as they have zero or vanishingly small magnetization. However, recent theoretical studies have proposed that such spin-axis change can be observed in AFMs<sup>24–26,29</sup>—for instance, through anisotropic magnetoresistance (AMR) effects<sup>25</sup>. Experimental demonstrations have been performed at room temperature in which rotation of antiferromagnetic moments was detected as the AMR changed by a few tens of milliohms, which is of the order of 0.1% of the total resistance<sup>28</sup>.

The AHE provides another useful probe for the spin-axis switch, and therefore may serve as an electrical means for reading magnetically stored information. In fact, in  $\text{Mn}_3\text{Sn}$  the Hall voltage can be readily detected as it generates a sizeable resistance jump at room temperature (for example,  $> 500$  m $\Omega$  for a 100-nm-thick thin film and  $|\Delta\rho_H/\rho| > 1\%$ ) and this material has no magnetoresistance up to several tesla. Thus, the sharp Hall resistance change can be easily tuned to be more than 10% of the total resistance by reducing the misalignment of Hall voltage contacts.

Whereas the remanent magnetization and thus stray fields of  $\text{Mn}_3\text{Sn}$  should be two to three orders of magnitude smaller than in ordinary FMs, its coercivity of a few hundred oersted is close to that of FMs used in magnetic devices. It will therefore be interesting to explore the possibility of using electrical means not only for reading but also for writing information—for example, by spin transfer torque<sup>24,26,29</sup>. Finally, we note that the present exceptionally large AHE found in an AFM with vanishingly small magnetization indicates that a large fictitious field due to Berry phase must exist in momentum space, and is expected to generate various effects including orbital ferromagnetism<sup>4,20</sup> and the spin Hall effect<sup>30</sup>. Exploration of such effects and their external-field control are suitable subjects for future studies.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 May; accepted 1 September 2015.**

**Published online 28 October 2015.**

- Nagaosa, N., Sinova, J., Onoda, S., MacDonald, A. H. & Ong, N. P. Anomalous Hall effect. *Rev. Mod. Phys.* **82**, 1539–1592 (2010).
- Shindou, R. & Nagaosa, N. Orbital ferromagnetism and anomalous Hall effect in antiferromagnets on the distorted fcc lattice. *Phys. Rev. Lett.* **87**, 116801 (2001).
- Metalidis, G. & Bruno, P. Topological Hall effect studied in simple models. *Phys. Rev. B* **74**, 045327 (2006).
- Martin, I. & Batista, C. D. Itinerant electron-driven chiral magnetic ordering and spontaneous quantum Hall effect in triangular lattice models. *Phys. Rev. Lett.* **101**, 156402 (2008).
- Chen, H., Niu, Q. & MacDonald, A. H. Anomalous Hall effect arising from noncollinear antiferromagnetism. *Phys. Rev. Lett.* **112**, 017205 (2014).
- Kübler, J. & Felser, C. Non-collinear antiferromagnets and the anomalous Hall effect. *Europhys. Lett.* **108**, 67001 (2014).
- Machida, Y., Nakatsuji, S., Onoda, S., Tayama, T. & Sakakibara, T. Time-reversal symmetry breaking and spontaneous Hall effect without magnetic dipole order. *Nature* **463**, 210–213 (2010).
- Tomiyoshi, S. & Yamaguchi, Y. Magnetic structure and weak ferromagnetism of  $\text{Mn}_3\text{Sn}$  studied by polarized neutron diffraction. *J. Phys. Soc. Jpn* **51**, 2478–2486 (1982).
- Brown, P. J., Nunez, V., Tasset, F., Forsyth, J. B. & Radhakrishna, P. Determination of the magnetic structure of  $\text{Mn}_3\text{Sn}$  using generalized neutron polarization analysis. *J. Phys. Condens. Matter* **2**, 9409–9422 (1990).
- Nagamiya, T., Tomiyoshi, S. & Yamaguchi, Y. Triangular spin configuration and weak ferromagnetism of  $\text{Mn}_3\text{Sn}$  and  $\text{Mn}_3\text{Ge}$ . *Solid State Commun.* **42**, 385–388 (1982).
- Krén, E., Paitz, J., Zimmer, G. & Zsoldos, É. Study of the magnetic phase transformation in the  $\text{Mn}_3\text{Sn}$  phase. *Physica B* **80**, 226–230 (1975).
- Ohmori, H., Tomiyoshi, S., Yamauchi, H. & Yamamoto, H. Spin structure and weak ferromagnetism of  $\text{Mn}_3\text{Sn}$ . *J. Magn. Magn. Mater.* **70**, 249–251 (1987).
- Tomiyoshi, S., Abe, S., Yamaguchi, Y., Yamauchi, H. & Yamamoto, H. Triangular spin structure and weak ferromagnetism of  $\text{Mn}_3\text{Sn}$  at low temperature. *J. Magn. Magn. Mater.* **54–57**, 1001–1002 (1986).
- Feng, W. J. *et al.* Glassy ferromagnetism in  $\text{Ni}_3\text{Sn}$ -type  $\text{Mn}_{3.1}\text{Sn}_{0.9}$ . *Phys. Rev. B* **73**, 205105 (2006).
- Miyasato, T. *et al.* Crossover behavior of the anomalous Hall effect and anomalous Nernst effect in itinerant ferromagnets. *Phys. Rev. Lett.* **99**, 086602 (2007).
- Neubauer, A., Pfleiderer, C., Ritz, R., Niklowitz, P. G. & Böni, P. Hall effect and magnetoresistance in  $\text{MnSi}$ . *Physica B* **404**, 3163–3166 (2009).
- Manyala, N. *et al.* Large anomalous Hall effect in a silicon-based magnetic semiconductor. *Nature Mater.* **3**, 255–262 (2004).
- Xiao, D., Chang, M.-C. & Niu, Q. Berry phase effects on electronic properties. *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
- Wan, X., Turner, A. M., Vishwanath, A. & Savrasov, S. Y. Topological semimetal and Fermi-arc surface states in the electronic structure of pyrochlore iridates. *Phys. Rev. B* **83**, 205101 (2011).
- Burkov, A. A. & Balents, L. Weyl semimetal in a topological insulator multilayer. *Phys. Rev. Lett.* **107**, 127205 (2011).
- Chappert, C., Fert, A. & Nguyen Van Dau, F. The emergence of spin electronics in data storage. *Nature Mater.* **6**, 813–823 (2007).
- Núñez, A. S., Duine, R. A., Haney, P. & MacDonald, A. H. Theory of spin torques and giant magnetoresistance in antiferromagnetic metals. *Phys. Rev. B* **73**, 214426 (2006).
- Shick, A. B., Khmelevskiy, S., Mryasov, O. N., Wunderlich, J. & Jungwirth, T. Spin-orbit coupling induced anisotropy effects in bimetallic antiferromagnets: a route towards antiferromagnetic spintronics. *Phys. Rev. B* **81**, 212409 (2010).
- MacDonald, A. H. & Tsoi, M. Antiferromagnetic metal spintronics. *Phil. Trans. R. Soc. A* **369**, 3098–3114 (2011).
- Park, B. G. *et al.* A spin-valve-like magnetoresistance of an antiferromagnet-based tunnel junction. *Nature Mater.* **10**, 347–351 (2011).
- Marti, X. *et al.* Room-temperature antiferromagnetic memory resistor. *Nature Mater.* **13**, 367–374 (2014).
- Gomonay, E. V. & Loktev, V. M. Spintronics of antiferromagnetic systems. *Low Temp. Phys.* **40**, 17–35 (2014).
- Hirsch, J. E. Spin Hall effect. *Phys. Rev. Lett.* **83**, 1834–1837 (1999).

**Acknowledgements** We thank M. Ikhlās and A. Nevidomskyy for discussions. This work was partially supported by PRESTO, by the Japan Science and Technology Agency, by Grants-in-Aid for Scientific Research (no. 25707030) and the Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers (no. R2604), and by Grants-in-Aid for Scientific Research on Innovative Areas (15H05882, 15H05883) from the Japanese Society for the Promotion of Science. The use of the facilities of the Materials Design and Characterization Laboratory at the Institute for Solid State Physics, The University of Tokyo, is acknowledged.

**Author Contributions** S.N. planned the experimental project, and S.N. and N.K. performed experiments and collected data. S.N., N.K. and T.H. wrote the paper and prepared figures; all authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.N. ([satoru@issp.u-tokyo.ac.jp](mailto:satoru@issp.u-tokyo.ac.jp)).

- Hall, E. H. On the “rotational coefficient” in nickel and cobalt. *Proc. Phys. Soc. Lond.* **4**, 325–342 (1880).
- Chien, C. L. & Westgate, C. R. *The Hall Effect and its Applications* (Plenum, 1980).

## METHODS

Polycrystalline samples were prepared by arc-melting mixtures of manganese and tin in a purified argon atmosphere. Excess manganese (10 mol%) over the stoichiometric amount was added to compensate for loss during arc-melting and single-crystal growth. The polycrystalline materials obtained were used to grow single crystals by the Czochralski method, using a commercial tetra-arc furnace (TAC-5100, GES). Our single-crystal and powder X-ray measurements indicate a single phase of hexagonal  $\text{Mn}_3\text{Sn}$  with lattice constants of  $a = 5.66(1)$  Å and  $c = 4.53(1)$  Å. Our scanning electron microscopy-energy dispersive X-ray analysis confirmed that  $\text{Mn}_3\text{Sn}$  is the bulk phase, and analysis using inductively coupled plasma spectroscopy found that the composition of the single crystals is  $\text{Mn}_{3.02}\text{Sn}_{0.98}$ .

For measurements of resistivity and magnetization, we used as-grown single crystals fashioned into bar-shaped samples after alignment using a Laue diffractometer. For our crystals, no annealing was performed below 800 °C, and thus the triangular spin configuration remains stable at least down to 50 K, below which the spins cant slightly towards the  $c$  axis<sup>11,14,15</sup>.

Magnetization measurements were made using a commercial SQUID magnetometer (MPMS, Quantum Design). Both longitudinal and Hall resistivities were measured by a standard four-probe method using a commercial system (PPMS, Quantum Design). The field dependence of the Hall resistivity was obtained after subtracting the longitudinal resistivity component, which is found to be constant as a function of the magnetic field.

The zero-field component of the AHE shown in Fig. 4 was estimated by the following method. We cooled samples from 400 K down to 5 K under a field of  $B_{\text{FC}} = 7$  T (−7 T), and subsequently at 5 K we decreased the field  $B$  down to +0 T (−0 T) without changing the sign of  $B$ . (Here we use +0 and −0 to indicate zero approached from the positive side and negative side, respectively.) Then, we measured the Hall voltage  $V_{\text{H}}(B \rightarrow +0)(V_{\text{H}}(B \rightarrow -0))$  in zero field after heating to various temperatures, stabilizing the temperature at each point. To remove the longitudinal resistance component due to the misalignment of the Hall voltage contacts, we estimated the zero-field component of the Hall resistance as  $R_{\text{H}}(B = 0) = [V_{\text{H}}(B \rightarrow +0) - V_{\text{H}}(B \rightarrow -0)]/2I$ , where  $I$  is the electric current. Three different samples were used for the three different field-cooling configurations shown in Fig. 4. The longitudinal resistivity at zero field  $\rho(B = 0)$  was measured concomitantly in the same procedures as those used for the Hall resistivity measurements. The zero-field remanent magnetization  $M(B = 0)$  was also measured using the same field-cooling procedures as used in both longitudinal and Hall resistivity measurements.

As discussed in the main text, the Hall resistivity in  $\text{Mn}_3\text{Sn}$  can be described by equation (1). The unconventional AHE term  $\rho_{\text{H}}^{\text{AF}}$  due to the non-collinear antiferromagnetic spin structure appears only in the  $a$ – $b$  plane components, and is absent when  $B \parallel c$ . Thus, to estimate the ordinary Hall coefficient, we measured the temperature dependence of the  $c$ -axis components of both the Hall resistivity  $\rho_{\text{H}}$  and the susceptibility  $\chi = M/B$  under  $B = 0.1$  T  $\parallel c$  (Extended Data Fig. 2a, b). Extended Data Figure 2c shows the Hall coefficient  $R_{\text{H}} = \rho_{\text{H}}/B$  versus the susceptibility  $M/B$ . A good fit using equation (1) with  $\rho_{\text{H}}^{\text{AF}} = 0$  was obtained for a wide range of temperature between 50 K and 300 K. The fitting yields  $R_0 = 3.0 \times 10^{-4} \text{ cm}^3 \text{ C}^{-1}$ , which corresponds to a carrier concentration of  $1.9 \times 10^{22} \text{ cm}^{-3}$ .  $R_5 = 0.86 \text{ cm}^3 \text{ C}^{-1}$  obtained by the fitting is consistent with another estimate at 300 K of  $R_5 = 0.5 \text{ cm}^3 \text{ C}^{-1}$ , which is obtained by using the slope of the isothermal curve of  $\rho_{\text{H}}$  versus  $M$  in  $B \parallel c$  (Fig. 2e). As  $|R_{\text{S}}| \gg |R_0|$ , the AHE dominates the Hall resistivity  $\rho_{\text{H}}$ , and the ordinary Hall contribution is negligibly small.

In magnetic conductors with relatively large resistivity, the AHE is dominated by contributions proportional to the resistivity squared,  $\rho^2$ , and thus it is useful to compare values of  $S_{\text{H}} = \mu_0 R_{\text{S}}/\rho^2$  (ref. 3) obtained from different magnetic materials. Extended Data Table 1 shows a list of magnets that show an AHE whose dominant contribution is proportional to  $\rho^2$ , including Fe, Ni, the helimagnet  $\text{MnSi}$  and the dilute magnetic semiconductor  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$  (refs 18, 19, 31–35). For comparison, we include in the list the non-collinear AFM  $\text{Mn}_5\text{Si}_3$  (ref. 36).

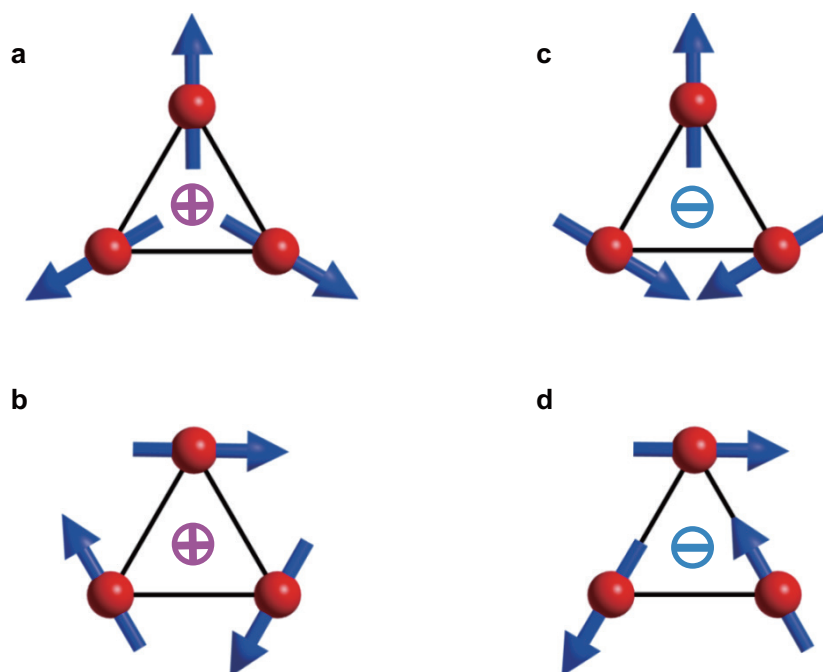
$S_{\text{H}}^0 = -\sigma_{\text{H}}(B = 0) / M(B = 0)$  estimated at zero field for  $\text{Mn}_3\text{Sn}$  using the isothermal curves in Figs 2 and 3 reaches  $-8.3 \text{ V}^{-1}$  (at 100 K), which is several orders of magnitude larger than in FMs such as Fe, Ni and  $\text{MnSi}$  (Extended Data Table 1). It is also much larger than the value for the non-collinear AFM  $\text{Mn}_5\text{Si}_3$  with  $T_{\text{N}} = 66$  K, which exhibits an AHE only by application of a magnetic field and thus has  $\rho_{\text{H}}^{\text{AF}} = 0$  (ref. 36). To further estimate the temperature dependence of  $S_{\text{H}}^0$ , we measured the remanent magnetization  $M(B = 0)$  using the same FC procedures as used for the estimate of  $\sigma_{\text{H}}(B = 0)$ . Figure 4b shows the temperature dependence of  $S_{\text{H}}^0 = -\sigma_{\text{H}}(B = 0) / M(B = 0)$  for  $\text{Mn}_3\text{Sn}$ . Again in comparison with  $S_{\text{H}}$  known for metallic FMs (Extended Data Table 1),  $S_{\text{H}}^0 \approx \rho_{\text{H}}^{\text{AF}}(B = 0) / [\rho^2(B = 0)M(B = 0)]$  is larger by a factor of  $>10$  and reaches  $-14 \text{ V}^{-1}$  at 100 K.

On the other hand, in a magnetic field, both the in-plane and  $c$ -axis components of  $\rho_{\text{H}}$  have the  $M$ -linear AHE contribution, namely,  $S_{\text{H}}\rho^2 M$ , as shown in Fig. 2e (broken lines). As  $\rho(B)$  remains constant with field,  $S_{\text{H}}$  estimated using this field-induced  $M$ -linear contribution is also constant with field: the  $S_{\text{H}}$  we obtain is much smaller than, and has a different sign from,  $S_{\text{H}}^0$ . Interestingly, they are found to be roughly isotropic, and have the same order of magnitude as found in itinerant FMs (Extended Data Table 1). The fact that  $S_{\text{H}}^0$  has a much larger value than, and a different sign from,  $S_{\text{H}}$  indicates that  $\rho_{\text{H}}^{\text{AF}}$  has a different origin from the conventional AHE, which is proportional to the magnetization.

To estimate  $\rho_{\text{H}}^{\text{AF}}$ , we subtracted the ordinary Hall effect  $R_0 B$  and the  $M$ -linear term  $S_{\text{H}}\rho^2 M$  from the isothermal curve of  $\rho_{\text{H}}$  by using the above estimates of both  $R_0$  and  $S_{\text{H}}$ . Note that the hysteresis in  $M(B)$  was also taken into account for this subtraction. Interestingly, the isothermal curve of  $\rho_{\text{H}}^{\text{AF}}$  is nearly independent of either  $B$  or  $M$ , and shows a clear sign change with a hysteresis when the staggered moment direction of the non-collinear spin structure is switched by rotating the small net ferromagnetic moment<sup>10,12</sup> (Fig. 2f, Extended Data Fig. 3). This indicates that the large zero-field AHE,  $\rho_{\text{H}}^{\text{AF}}$ , comes from the non-collinear antiferromagnetism and is most probably driven by its Berry curvature effects.

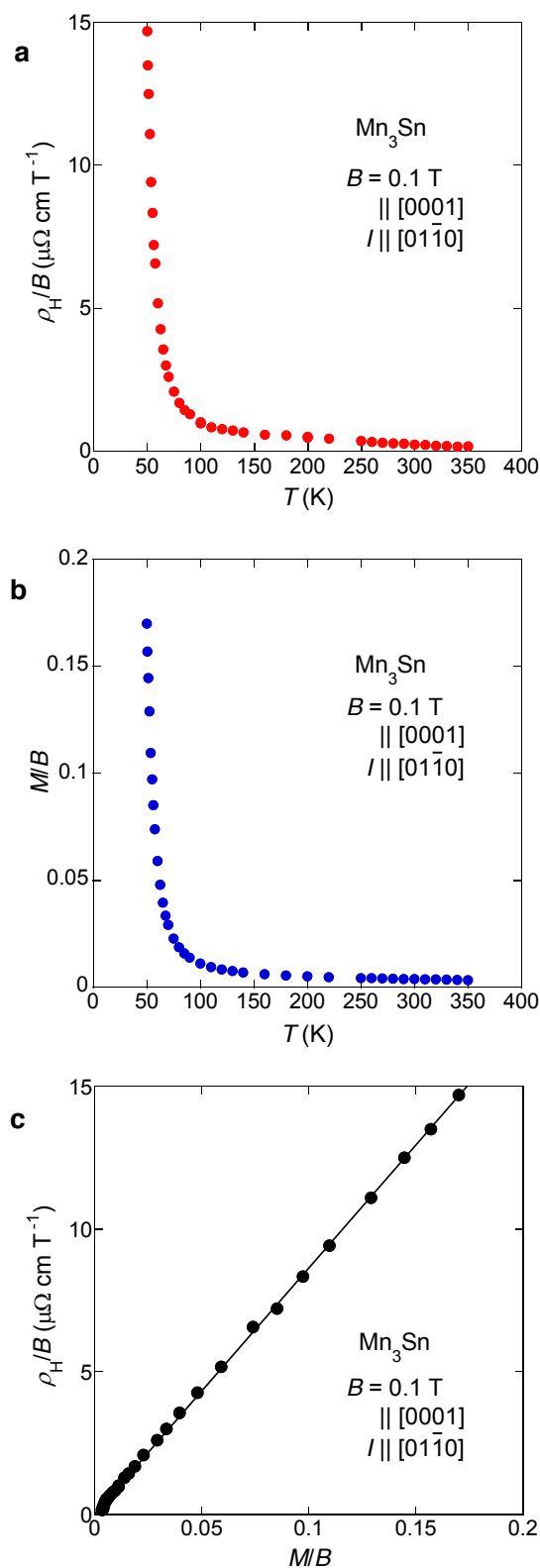
- Dheer, P. N. Galvanomagnetic effects in iron whiskers. *Phys. Rev.* **156**, 637–644 (1967).
- Jan, J.-P. & Gijsman, H. M. L'effet Hall du fer et du nickel aux basses températures. *Physica* **18**, 339–355 (1952).
- Husmann, A. & Singh, L. J. Temperature dependence of the anomalous Hall conductivity in the Heusler alloy  $\text{Co}_2\text{CrAl}$ . *Phys. Rev. B* **73**, 172417 (2006).
- Chun, S. H. *et al.* Interplay between carrier and impurity concentrations in annealed  $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ : intrinsic anomalous Hall effect. *Phys. Rev. Lett.* **98**, 026601 (2007).
- Onose, Y. & Tokura, Y. Doping dependence of the anomalous Hall effect in  $\text{La}_{1-x}\text{Sr}_x\text{CoO}_3$ . *Phys. Rev. B* **73**, 174421 (2006).
- Sürgers, C., Fischer, G., Winkel, P. & v. Löhneysen, H. Large topological Hall effect in the non-collinear phase of an antiferromagnet. *Nature Commun.* **5**, 3400 (2014).



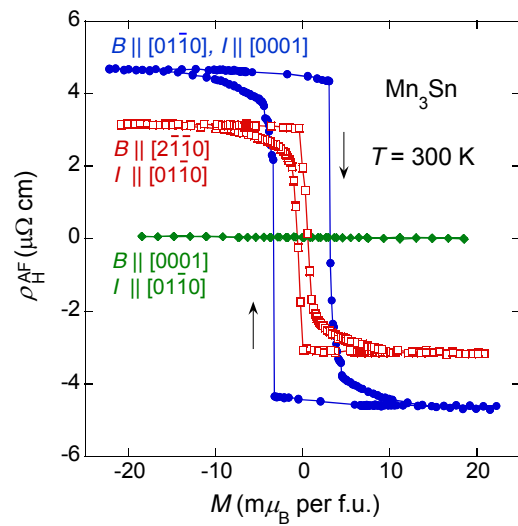


**Extended Data Figure 1 | Normal and inverse triangular spin structures.** **a–d**, Examples of normal (**a**, **b**) and inverse (**c**, **d**) triangular spin structures. An inverse triangular spin structure has the opposite sign of the vector

spin chirality to a normal one. As each Mn moment has the local easy-axis parallel to the direction towards its in-plane nearest-neighbour Sn sites, the case shown in **d** is realized in  $\text{Mn}_3\text{Sn}$  (refs 10–12).



**Extended Data Figure 2 | Estimation of the ordinary Hall effect using the *c*-axis components of the Hall resistivity.** **a, b,** Temperature dependence of the Hall resistivity divided by  $B$ ,  $\rho_H/B$  (**a**), and the susceptibility  $M/B$  obtained under a field of 0.1 T along the *c* axis (**b**). Measurements were made above  $T = 50$  K, where no spontaneous components were observed. **c,** Plot of  $\rho_H/B$  versus  $M/B$ ; here the temperature is an implicit parameter. The solid line indicates a linear fit to equation (1), with  $\rho_H^{\text{AF}} = 0$  defined in Methods.



**Extended Data Figure 3 | Magnetization dependence of the Hall effect driven by the antiferromagnetic order.** Figure shows anisotropic isothermal curves of  $\rho_H^{\text{AF}}$  as a function of  $M$  at 300 K.



Extended Data Table 1 | Hall effect parameters for ferromagnets and antiferromagnets

Material	$T$ [K]	$n$ [ $10^{22} \text{ cm}^{-3}$ ]	$S_H^0$ [ $\text{V}^{-1}$ ]	$S_H$ [ $\text{V}^{-1}$ ]	Reference
Mn <sub>3</sub> Sn (AFM)					
$B \parallel [01\bar{1}0], I \parallel [0001]$	100	+1.9	−8.3	0.24	This work
$B \parallel [01\bar{1}0], I \parallel [0001]$	300	+1.9	−2.2	0.069	This work
$B \parallel [2\bar{1}\bar{1}0], I \parallel [01\bar{1}0]$	300	+1.9	−3.6	0.085	This work
$B \parallel [0001], I \parallel [01\bar{1}0]$	300	+1.9		0.1	This work
Fe(bcc) (FM) ( $B \parallel [100]$ )	298	+34		0.06	Ref. 31
Ni (FM)	293	−11		−0.14	Ref. 32
MnSi (FM)	$\leq 29.5$	+5.9 – +8.5		−0.19	Ref. 18
Fe <sub>1−<math>x</math></sub> Co <sub><math>x</math></sub> Si (FM)	$\leq 53$	−0.26 – −0.10		0.022	Ref. 19
Co <sub>2</sub> CrAl (FM)	36 – 278	> 0.03		0.039	Ref. 33
Ga <sub>1−<math>x</math></sub> Mn <sub><math>x</math></sub> As (FM)	15	+0.01 – +0.09		0.04 – 0.06	Ref. 34
La <sub>0.7</sub> Sr <sub>0.3</sub> CoO <sub>3</sub> (FM)	2	+6		0.06	Ref. 35
Mn <sub>5</sub> Si <sub>3</sub> C <sub>0.8</sub> (FM)	100	+3		0.013	Ref. 36
Mn <sub>5</sub> Si <sub>3</sub> (AFM)	20	+0.3		−0.009	Ref. 36

Temperature ( $T$ ), effective carrier density ( $n$ ) and the coefficient  $S_H$  of the AHE are shown for various magnets. Here  $S_H$  is mostly determined by using the isothermal data of the Hall resistivity, magnetoresistivity and magnetization. For Mn<sub>3</sub>Sn, two types of  $S_H$  are shown, namely,  $S_H$  for the AHE component that is linear in  $M$ , and  $S_H^0$  for the zero-field spontaneous component of  $\rho_H$  (Methods). The sign of  $n$  indicates the type of carriers, namely, hole (+) or electron (−). Field ( $B$ ) and current ( $I$ ) direction are specified for some cases. AFM, antiferromagnet; FM, ferromagnet; bcc, body-centred cubic. Data are from this work, and refs 18, 19, 31–36.

# Liquids with permanent porosity

Nicola Giri<sup>1</sup>, Mario G. Del Pópolo<sup>2,3</sup>, Gavin Melaugh<sup>2</sup>, Rebecca L. Greenaway<sup>4</sup>, Klaus Rätzke<sup>5</sup>, Tönjes Koschine<sup>5</sup>, Laure Pison<sup>6</sup>, Margarida F. Costa Gomes<sup>6</sup>, Andrew I. Cooper<sup>4</sup> & Stuart L. James<sup>1</sup>

Porous solids such as zeolites<sup>1</sup> and metal–organic frameworks<sup>2,3</sup> are useful in molecular separation and in catalysis, but their solid nature can impose limitations. For example, liquid solvents, rather than porous solids, are the most mature technology for post-combustion capture of carbon dioxide because liquid circulation systems are more easily retrofitted to existing plants. Solid porous adsorbents offer major benefits, such as lower energy penalties in adsorption–desorption cycles<sup>4</sup>, but they are difficult to implement in conventional flow processes. Materials that combine the properties of fluidity and permanent porosity could therefore offer technological advantages, but permanent porosity is not associated with conventional liquids<sup>5</sup>. Here we report free-flowing liquids whose bulk properties are determined by their permanent porosity. To achieve this, we designed cage molecules<sup>6,7</sup> that provide a well-defined pore space and that are highly soluble in solvents whose molecules are too large to enter the pores. The concentration of unoccupied cages can thus be around 500 times greater than in other molecular solutions that contain cavities<sup>8–10</sup>, resulting in a marked change in bulk properties, such as an eightfold increase in the solubility of methane gas. Our results provide the basis for development of a new class of functional porous materials for chemical processes, and we present a one-step, multigram scale-up route for highly soluble ‘scrambled’ porous cages prepared from a mixture of commercially available reagents. The unifying design principle for these materials is the avoidance of functional groups that can penetrate into the molecular cage cavities.

The structural rigidity and robustness of solids allows them to contain permanent, uniform cavities of precise size and shape. By contrast, liquids have fluid structures, and any ‘porosity’ is limited to poorly defined and transient intermolecular cavities<sup>11</sup>, most of which are smaller than typical molecules. The scaled particle theory<sup>12,13</sup> predicts that the solubility of a solute in a liquid is primarily influenced by the work required to generate the cavities that accommodate the solute. Thus, if high concentrations of permanent, molecule-sized cavities could be created in a liquid, the solvating and solute-transporting characteristics of that liquid should be strongly affected<sup>5,14,15</sup>. For example, solutes might become much more soluble because no work would be required to create cavities<sup>12,13</sup>. This effect should be specific to solute molecules of complementary size and shape to the pores, as is common for selective adsorption in porous solids.

We have prepared ‘porous liquids’ by taking rigid organic cage molecules, each of which defines a molecular pore space, and dissolving them at high concentration in a solvent that is too large to enter the pores (Fig. 1). Hence, the pores in the cages remain empty and available to solutes. In our first system, a crown ether, 15-crown-5, was chosen as the solvent because it is a liquid at room temperature and because it consists of large molecules with low surface curvature. Hence, no part of any solvent molecule can fit into the cage pores. Into this solvent were dissolved rigid organic cages. Each cage has a cavity of ~5 Å diameter at its centre that is accessible through four access windows of ~4 Å in

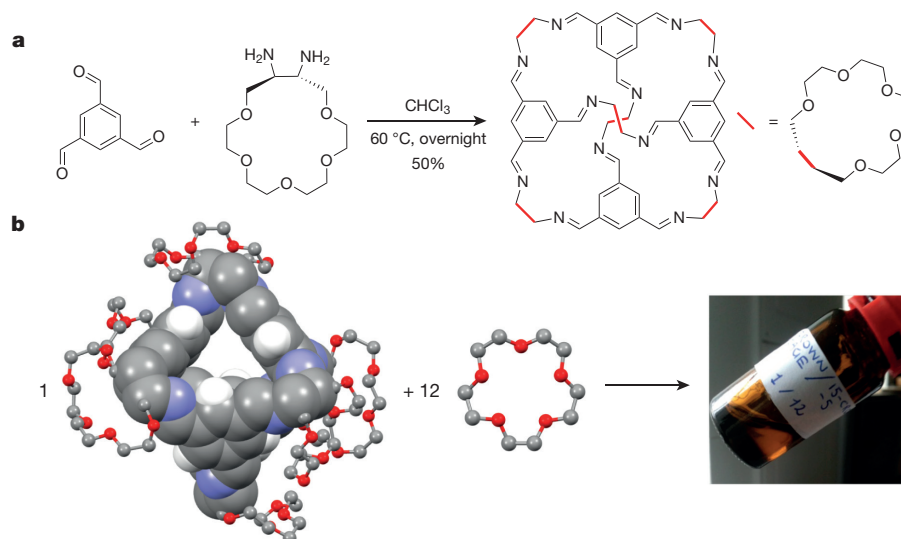
diameter. Structurally similar but much less soluble cages are known to absorb small gas molecules into their pores in the solid state<sup>6,7</sup>. To achieve high solubility in the bulky 15-crown-5 solvent, each cage was functionalized on the outside with six crown ether groups by using a crown-ether functionalized diamine to prepare the cage. As a result, we could prepare, at room temperature, an extremely concentrated liquid phase (44 wt%) where just 12 solvent molecules are needed to dissolve each cage molecule. More typical solutions of molecules with cavities contain many hundreds or thousands of solvent molecules for each cavity. This liquid flows at room temperature, despite the high concentration of cages. The viscosity of the liquid was 20 cP to >140 cP in the range 298 K to 323 K, depending on the cage concentration (Supplementary Figs 10 and 11). In contrast, the viscosity of pure 15-crown-5 is 8.9 cP to 22.4 cP over this temperature range. Organic cages on their own can also form liquid phases close to room temperature if they have long alkyl chain substituents, but in those cases the chains occupy the cage cavities, thus removing the porosity<sup>15</sup>.

Molecular dynamics (MD) simulations gave detailed insight into the structure of this ‘porous liquid’. During 100-ns simulations, 100% of the cages remained empty at all times, both at 350 K and 400 K. Further modelling confirmed that occupation of the cages by the 15-crown-5 solvent molecules, or by the crown-ether substituents on the outside of the cages, would incur high free energy costs (full details are given in Supplementary Information). In fact, the free energy computed along a reaction coordinate that eventually forces the occupation of the cage by a 15-crown-5 solvent molecule shows no minimum when the solvent is inside the cage (Supplementary Fig. 14). Therefore, we conclude that the cages remain empty in the porous liquid over timescales longer than the MD simulations.

The distribution of cavity sizes,  $\rho(R)$ , was computed for the porous liquid,  $\rho_{\text{mix}}(R)$ , and for the pure 15-crown-5 solvent,  $\rho_{\text{sol}}(R)$ , both at the same temperature and pressure. Here  $\rho(R)$  is the so-called insertion probability, which measures the likelihood that a hard sphere of radius  $R$  could be inserted at any arbitrary point within the liquid without overlapping with the van der Waals volume of any atom of the liquid<sup>11</sup>. The snapshot in Fig. 2a represents a typical configuration in the MD simulation for the 1:12 mixture at 350 K. The coloured surfaces show the empty cavities both inside (red) and outside (yellow) the cage cores that are accessible to spherical probes of radius 2.0–2.6 Å. This probe radius range was chosen because it reveals the cavities that are sufficiently large to accommodate methane as a solute molecule. Figure 2a shows that the cage cavities account for most of the large, methane-sized cavities in the liquid. This can be quantified with respect to the pure 15-crown-5 solvent by calculating the relative porosity of the mixture, defined as  $V_{\text{rel}}(R) = \rho_{\text{mix}}(R)/\rho_{\text{sol}}(R)$ . A value of  $V_{\text{rel}}(R)$  greater than one indicates that the fractional free volume accessible to a spherical probe of radius  $R$  is larger in the porous liquid than in the pure solvent. As might be expected, the cages have no effect on the concentration of small, submolecular cavities in the liquid, and  $V_{\text{rel}}(R)$  is around 1 for  $R < 0.1$  nm (Fig. 2b). However, the

<sup>1</sup>School of Chemistry and Chemical Engineering, Queen’s University Belfast, David Keir Building, Stranmillis Road, Belfast BT9 5AG, UK. <sup>2</sup>School of Mathematics and Physics, Queen’s University Belfast, University Road, Belfast BT7 1NN, UK. <sup>3</sup>CONICET and Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, Padre Jorge Contreras 1300, CP5500 Mendoza, Argentina.

<sup>4</sup>University of Liverpool, Department of Chemistry and Centre for Materials Discovery, Crown Street, Liverpool L69 7ZD, UK. <sup>5</sup>Technische Fakultät der Universität Kiel, Kaiserstrasse 2, D-24143 Kiel, Germany. <sup>6</sup>Institut de Chimie de Clermont-Ferrand, UMR 6296, CNRS, Université Blaise Pascal, 24 avenue Blaise Pascal, 63178 Aubière, France.



**Figure 1 | Preparation of the porous liquid.** **a**, Synthesis of the crown-ether cage. **b**, The empty, highly soluble cage molecule, left, defines the pore space; the 15-crown-5 solvent, middle, provides fluidity but cannot enter the cage cavities. The concentrated solution (porous liquid) flows at room temperature, right. Key: C, grey; O, red; N, blue; H, white.

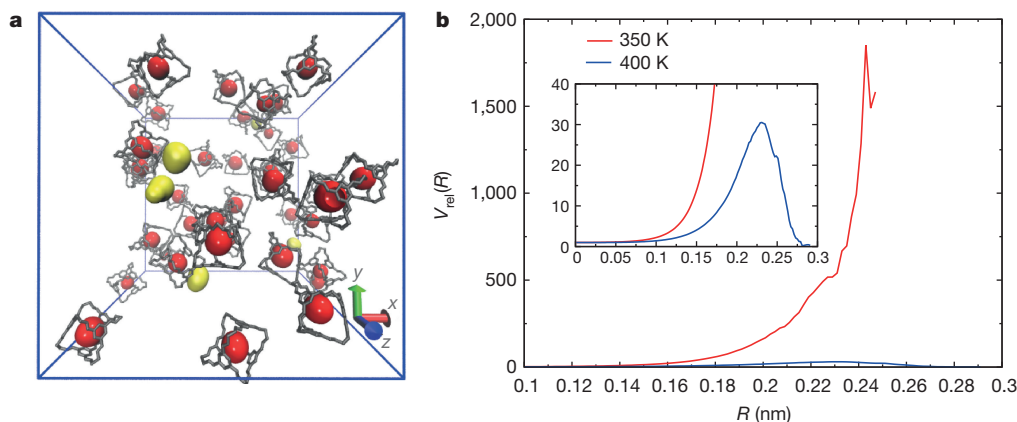
cages dramatically increase the concentration of cavities with radii in the range 0.1–0.25 nm. This effect is particularly strong at 350 K, where  $V_{\text{rel}}(R)$  approaches 1,900 for  $R = 0.24$  nm.  $V_{\text{rel}}(R)$  is larger at 350 K than at 400 K because more large cavities form transiently in the solvent at higher temperatures owing to increased thermal motion. Overall, the fractional void volume in the porous liquid due to the cages was estimated as 0.7% of the total volume. This is small compared to typical pore volumes in porous solids, but it can dramatically affect the solubility of solutes in the liquid.

Positron ( $e^+$ ) annihilation lifetime spectroscopy (PALS) experiments were used to investigate whether the cages in the porous liquid were empty. PALS probes the electron density distribution in materials, in particular the presence of empty pores. Exposure of an insulating material to a positron source such as  $^{22}\text{Na}$  leads to the formation of ortho-positronium (o-Ps; that is, an  $e^+$  and an  $e^-$  with parallel spins bound into an exotic atom) within the material. The lifetime of the positron (that is, the time between injection of the positron into the material and the decay of o-Ps) can be correlated to the average pore diameter in the material by a well-established model<sup>16</sup>, that is, the larger the pore the lower the decay probability and the longer the lifetime.

Space-filling rendering highlights the core of the cage. Ball and stick rendering represents the crown-ether substituents on the cage and the 15-crown-5 solvent. All H atoms except those attached to aromatic rings of the cage compound have been omitted for clarity.

Thus, by comparing the o-Ps lifetimes observed in the pure (solid) cage material, the pure 15-crown-5 solvent and the porous liquid, we investigated whether the cages in the porous liquid were empty. Several measurements were made on each material to check for reproducibility and to enable separation of the raw data for the porous liquid into constituent components. The o-Ps lifetime of  $2.05 \pm 0.1$  ns measured for the pure solid cage at 30 °C correlates to an average cavity diameter of 0.55 nm, in reasonable agreement with our MD simulations. The o-Ps lifetime of  $3.00 \pm 0.1$  ns for the pure 15-crown-5 solvent at 30 °C is also in agreement with previous literature<sup>17</sup>. This latter measurement can be explained by the known phenomenon of o-Ps bubble formation<sup>17</sup>, rather than by any pre-existing pores in the pure 15-crown-5 solvent. The o-Ps lifetime measured for the porous liquid at 30 °C was  $2.34 \pm 0.02$  ns. If the cages are empty in the porous liquid, then the o-Ps lifetime should correspond to a combination of the lifetimes measured for the pure cage and the pure 15-crown-5 solvent.

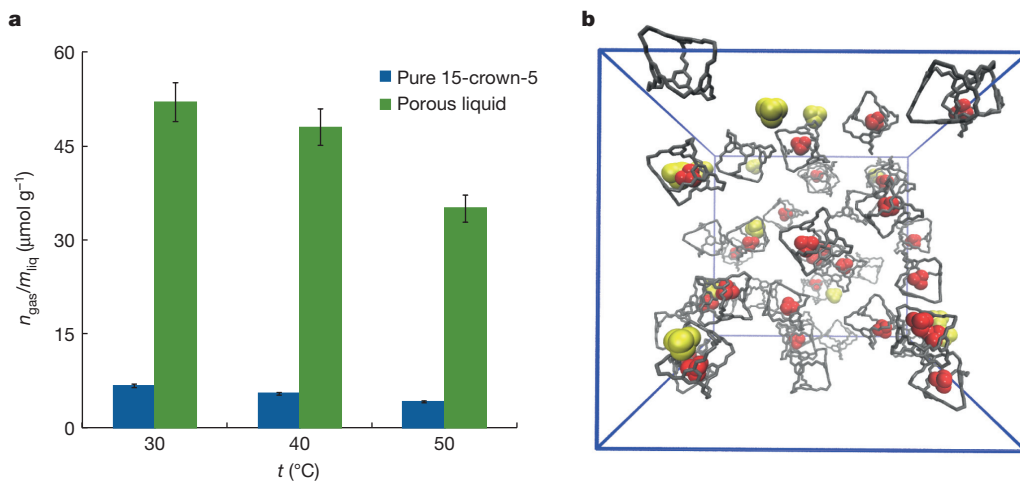
To investigate whether the o-Ps lifetime for the porous liquid could be separated into these individual contributions, both free and forced fits were applied to the raw data. Allowing both components to refine freely did not give a satisfactorily clear separation (large error bars were



**Figure 2 | Molecular simulations for the porous liquid show unoccupied molecular-sized pores.** **a**, Representative configuration of the porous liquid at 350 K. To highlight the cage cores, crown-ether solvent molecules and crown-ether substituents on the cages have been omitted. Red and yellow surfaces indicate empty pores inside or outside the cages,

respectively. **b**, Relative porosity,  $V_{\text{rel}}(R)$ , of the porous liquid at 350 K and 400 K. Inset, expansion of the 400 K result. At 350 K, the porous liquid has around 1,900 times as many methane-sized cavities (probe radius  $\sim 0.24$  nm) than does the pure solvent.





**Figure 3 | Dissolution of methane in the porous liquid.** **a**, Methane is around 8 times more soluble per mass of the porous liquid at 1 atm pressure than per mass of the pure 15-crown-5 solvent. Error bars represent overall uncertainty of experimental data calculated by error propagation of the s.d. of the measured quantities. **b**, Molecular simulation

observed; see Supplementary Fig. 23). However, a forced fit where one component was fixed at 3.00 ns, consistent with the anticipated contribution from the pure 15-crown-5 solvent, and where the other component was allowed to refine freely, did result in satisfactory separation of these two components (Supplementary Fig. 24). The freely refined component had a value of  $1.91 \pm 0.08$  ns, which corresponds well to the lifetime observed for the pure empty cage ( $2.05 \pm 0.1$  ns). The PALS measurements are thus consistent with the presence of empty cages, or molecular ‘pores’, in the porous liquid.

Both MD simulations and PALS data suggested that the porous liquid might have a dramatically increased capacity to dissolve solute molecules up to  $\sim 0.5$  nm in diameter relative to the pure crown solvent, particularly at temperatures of 350 K or lower (Fig. 2b). Acidic gases, such as carbon dioxide, are absorbed readily by conventional, non-porous solvents such as water or aqueous amines<sup>4</sup>. We therefore focused our initial studies on methane, the main component of natural gas. Methane, like other important gases such as hydrogen, lacks the Lewis acidity of carbon dioxide and hence cannot be absorbed using liquid amines. The methane solubility in pure 15-crown-5 at 30  $^{\circ}\text{C}$  was found to be  $6.7 \mu\text{mol g}^{-1}$ . By contrast, the solubility of methane in the porous liquid was  $52 \mu\text{mol g}^{-1}$  at 30  $^{\circ}\text{C}$ . This represents, approximately, an eightfold increase in solubility in the porous liquid compared to the pure solvent. The methane solubility decreased at higher temperatures, both in the porous liquid and in the non-porous solvent, but the solubility was much higher in the porous liquid at all temperatures studied (Fig. 3a).

Molecular simulations for solutions of methane in the porous liquid suggested that most methane molecules would reside in the cage cavities, rather than in transient intermolecular cavities (see Fig. 3b and Supplementary Information for details). For example, at 350 K and a gas pressure of 1 atm, 70% of the methane molecules in the liquid were located inside the cages, that is, less than 2.5 Å from a cage centre. These simulations support the methane solubility measurements by predicting that methane should be significantly more soluble in the porous liquid. The enhanced solubility cannot be ascribed to a simple solvating effect of the aromatic cage walls because methane solubility in pure aromatic solvents, such as benzene, is lower than for the porous liquid ( $26.8 \mu\text{mol g}^{-1}$  at 25  $^{\circ}\text{C}$ )<sup>18,19</sup>. The methane absorption capacity is remarkably high for a liquid, albeit much lower than for many porous solids, such as zeolitic imidazolate frameworks<sup>20</sup>.

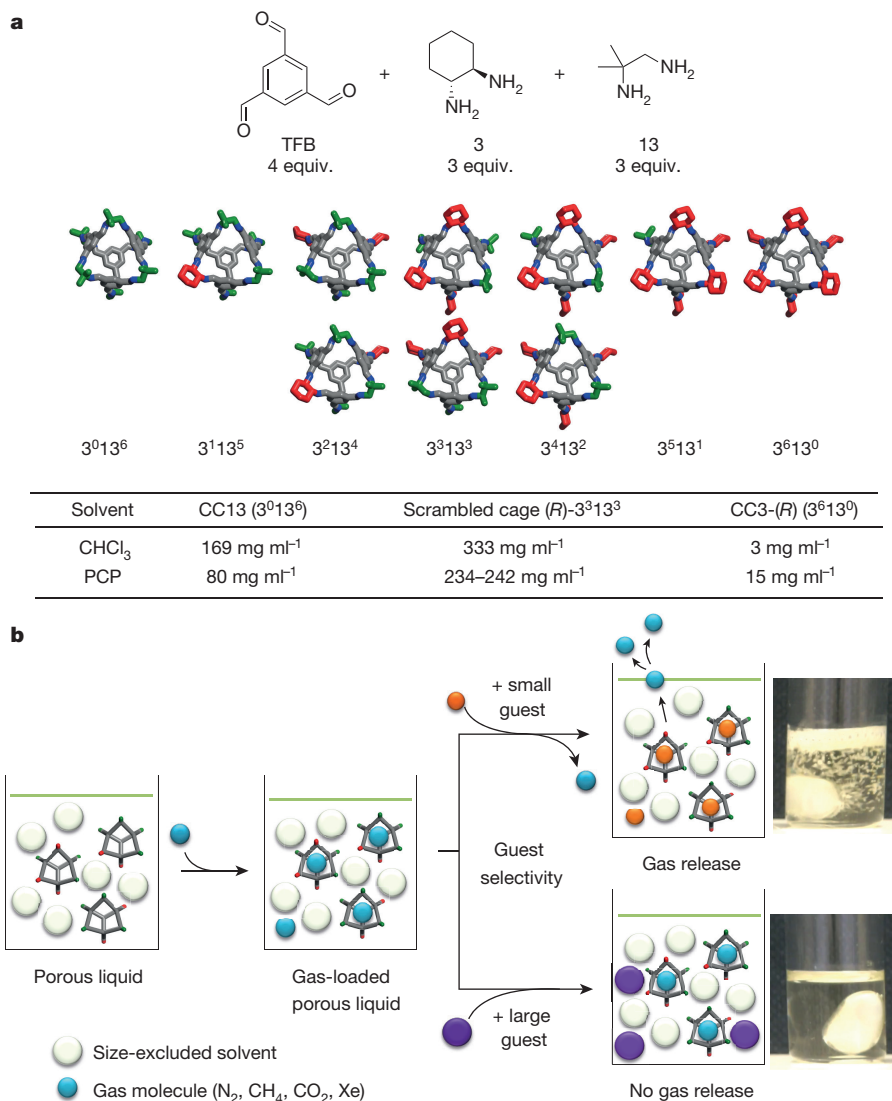
This porous molecular liquid differs fundamentally from porous macromolecular liquids containing hollow colloidal silica spheres<sup>21</sup>. Solutions of other types of organic cages with smaller internal pores,

of the porous liquid with dissolved methane (grand canonical Monte Carlo simulation at 350 K, 1 atm). Methane is shown in red when inside a cage core ( $< 2.5$  Å from the cage centre) or yellow when outside. Solvent and cage substituents are omitted for clarity.

known as carcerands and cryptophanes, also have pores that are most probably unsolvated because even relatively small solvent molecules are excluded<sup>8–10</sup>. The existence of permanent cavities in these carcer-and/cryptophane systems was not, however, supported by simulations or by PALS measurements, nor was enhanced bulk solubility of gases noted. This is probably because the solutions were very dilute; for example, based on a 2 mM host solution of a cryptophane in  $\text{CDCl}_3$ , as described in ref. 10, the molar ratio of cage to solvent molecules was 1:6,000, compared with the 1:12 solution that we study here. The cryptophane cavity volume in that study<sup>10</sup> was reported to be  $81 \text{ Å}^3$ . If the cryptophane is assumed not to have collapsed, then this represents just 0.01% of the total volume, which is 70 times lower than in our study. The much higher density of cavities in our system demonstrates a previously unexplored way to control the bulk solubility of gases and other solutes in liquids.

Although porous liquids of the type we report here might not be competitive with porous solids for gas storage, we envisage other applications, such as gas separation, which utilize the high concentration of prefabricated cavities in the liquid. A practical scale-up issue is the six-step cage synthesis (Supplementary Information), which has modest overall yield (3.1%–6.5%). Also, the viscosity of the crown-ether porous liquid is rather high (20 cP to  $> 140$  cP). We therefore developed an alternative, more scalable route, where a mixture of diamines is used to make a mixture of ‘scrambled’ cages<sup>22</sup> that have the same core cage dimensions as the crown-ether derivative discussed above. The resulting mixture of scrambled cage molecules has much higher solubility in common organic solvents than similar cages prepared from a single diamine (see table, Fig. 4a) because of increased structural disorder<sup>23</sup>.

Multi-gram quantities of the scrambled cage mixture are easily prepared from commercially available starting materials via this one-step synthesis (Fig. 4a; 77% yield). As for the crown-ether derivatized cage described above, this scrambled cage mixture can form concentrated solutions ( $> 10$  wt%) in a solvent, hexachloropropene (PCP), which like 15-crown-5 is size-excluded from the cage cavity. These solutions were at least ten times less viscous (11.7 cP at 295 K, 0.112 mmol of cage per 1 g of PCP) than the crown-ether-based porous liquid described above at comparable cage concentrations because of the lack of bulky crown-ether substituents. Like the porous liquid prepared from the crown-ether derivatized cage, this porous liquid prepared from the scrambled mixture of cages showed much higher methane gas solubility than the non-porous PCP solvent ( $51 \mu\text{mol g}^{-1}$  versus  $6.7 \mu\text{mol g}^{-1}$  at 293 K). Other gases such as nitrogen, carbon dioxide and xenon also showed enhanced solubilities in the scrambled-cage-based porous liquid, and



**Figure 4 | Porous liquids based on scrambled cages.** **a**, Reaction of the trialdehyde, TFB, with two different diamines, 3 and 13, bearing cyclohexane and dimethyl groups, respectively, gives rise to a statistical mixture of cages which differ from each other in the number and positions of their dimethyl (green) and cyclohexane (red) groups. Nomenclature  $3^x 13^y$  defines the number of cyclohexane groups ( $x$ ) and number of dimethyl groups ( $y$ ) for each cage. Three cages ( $3^2 13^4$ ,  $3^3 13^3$  and  $3^4 13^2$ ) each exist as a pair of isomers, as indicated by the additional structures shown in the second row of products. The average structure of this scrambled mixture is  $3^3 13^3$ . Mixtures of ‘scrambled’ cages were thus

prepared in a scalable, one-pot synthesis and are up to 100 times more soluble than cages prepared from single diamines, such as CC3-( $R$ ) (ref. 6) or CC13 (ref. 23). PCP is hexachloropropene solvent which is size-excluded from the cage cavities. **b**, The porous liquids show enhanced solubilities for methane, nitrogen, carbon dioxide and xenon (Supplementary Figs 32–35). Addition of a small, interpenetrating guest,  $\text{CHCl}_3$ , liberates xenon gas rapidly from a xenon-saturated porous liquid (top photograph; Supplementary Video 1) whereas addition of a large, size-excluded guest, 1-*t*-butyl-3,5-dimethylbenzene, does not (bottom photograph).

the trend in gas solubility follows the calculated Henry’s coefficients<sup>24</sup> and isosteric heats of sorption for these gases in solid porous cages of this size and geometry (Supplementary Fig. 35), again suggesting that gas solubility is dominated by the unoccupied cage cavities. A shift in the  $^1\text{H}$  NMR signal for methane was observed from  $-0.24$  p.p.m. (using tetramethylsilane, TMS, as standard) in neat PCP to  $-2.80$  p.p.m. in the scrambled-cage-based porous liquid (Supplementary Fig. 30). This strong shielding effect ( $\Delta\delta = -2.56$  p.p.m.) supports the presence of methane in the cage cavities of the scrambled porous liquid on the NMR timescale<sup>9,10</sup>.

As for the porous liquids prepared from crown-ether derivatized cages, the increased gas solubility in the scrambled-cage-based porous liquid stems from the absence of functional groups that can penetrate the cage cavities, in this case because the peripheral methyl groups on the scrambled cages are too small to do this. By contrast, other organic molecules of the correct size can compete

with gases for the cage cavities. For example, addition of chloroform rapidly displaces xenon from a sample of the scrambled-cage-based porous liquid that was saturated with xenon, whereas addition of a larger, size-excluded guest, 1-*t*-butyl-3,5-dimethylbenzene, does not (Fig. 4b; Supplementary Video 1). This allows dramatic solubility switching to occur with the addition of a remarkably small amount of a second, smaller solvent. For example, 10.05 ml STP of xenon is released from 5.74 g of the scrambled-cage-based porous liquid by addition of just 0.046 ml (1.29 vol.%) of  $\text{CHCl}_3$ . This remarkably sharp switching behaviour can be ascribed to a molecular displacement mechanism, rather than to a change in the bulk solvent properties.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 12 December 2014; accepted 12 October 2015.**

- Wright, P. A. *Microporous Framework Solids* (Royal Society of Chemistry, 2007).
- Cheetham, A. K., Férey, G. & Loiseau, T. Open-framework inorganic materials. *Angew. Chem. Int. Edn* **38**, 3268–3292 (1999).
- Kitagawa, S., Kitaura, R. & Noro, S. Functional porous coordination polymers. *Angew. Chem. Int. Edn* **43**, 2334–2375 (2004).
- D'Alessandro, D. M., Smit, B. & Long, J. R. Carbon dioxide capture: prospects for new materials. *Angew. Chem. Int. Edn* **49**, 6058–6082 (2010).
- O'Reilly, N., Giri, N. & James, S. L. Porous liquids. *Chem. Eur. J.* **13**, 3020–3025 (2007).
- Tozawa, T. *et al.* Porous organic cages. *Nature Mater.* **8**, 973–978 (2009).
- Jones, J. T. A. *et al.* Modular and predictable assembly of porous organic molecular crystals. *Nature* **474**, 367–371 (2011).
- Robbins, T. A., Knobler, C. B., Bellew, D. R. & Cram, D. J. A highly adaptive and strongly binding hemiacerand. *J. Am. Chem. Soc.* **116**, 111–122 (1994).
- Chaffee, K. E., Fogarty, H. A., Brotin, T., Goodson, B. M. & Dutasta, J.-P. Encapsulation of small gas molecules by cryptophane-111 in organic solution. 1. Size- and shape-selective complexation of simple hydrocarbons. *J. Phys. Chem. A* **113**, 13675–13684 (2009).
- Little, M. A. *et al.* Synthesis and methane-binding properties of disulfide-linked cryptophane-0.0.0. *Angew. Chem. Int. Edn* **51**, 764–766 (2012).
- Pohorille, A. & Pratt, L. R. Cavities in molecular liquids and the theory of hydrophobic solubilities. *J. Am. Chem. Soc.* **112**, 5066–5074 (1990).
- Pierotti, R. A. The solubility of gases in liquids. *J. Phys. Chem.* **67**, 1840–1845 (1963).
- Pierotti, R. A. A scaled particle theory of aqueous and non-aqueous solutions. *Chem. Rev.* **76**, 717–726 (1976).
- Giri, N. *et al.* Alkylated organic cages: from porous crystals to neat liquids. *Chem. Sci.* **3**, 2153–2157 (2012).
- Melaugh, G., Giri, N., Davidson, C. E., James, S. L. & Del Pópolo, M. G. Designing and understanding permanent microporosity in liquids. *Phys. Chem. Chem. Phys.* **16**, 9422–9431 (2014).
- Mogensen, O. E. (ed.) *Positron Annihilation in Chemistry* (Springer Series in Chemical Physics 58, Springer, 1995).
- Mahmood, T., Cheng, K. L. & Yean, Y. C. Microanalysis of open spaces in crown ethers by using a novel probe: positron annihilation spectroscopy. In *Third International Workshop on Positron and Positronium Chemistry* (ed. Jean, Y. C.) 640 (World Scientific, 1990).
- Lannung, A. & Gjaldbæk, J. C. The solubility of methane in hydrocarbons, alcohols, water and other solvents. *Acta Chem. Scand.* **14**, 1124–1128 (1960).
- Darwish, N. A., Gasem, K. A. M. & Robinson, R. L. Jr. Solubility of methane in benzene, naphthalene, phenanthrene and pyrene at temperatures from 323 to 433 K and pressures to 11.3 MPa. *J. Chem. Eng. Data* **39**, 781–784 (1994).
- Houndonougbo, Y. *et al.* A combined experimental–computational investigation of methane adsorption and selectivity in a series of isorecticular zeolitic imidazolate frameworks. *J. Phys. Chem. C* **117**, 10326–10335 (2013).
- Zhang, J. *et al.* Porous liquids: a promising class of media for gas separation. *Angew. Chem. Int. Edn* **54**, 932–936 (2015).
- Jiang, S. *et al.* Porous organic molecular solids by dynamic covalent scrambling. *Nature Commun.* **2**, 207 (2011).
- Hasell, T. *et al.* Controlling the crystallization of porous organic cages: molecular analogs of isorecticular frameworks using shape-specific directing solvents. *J. Am. Chem. Soc.* **136**, 1438–1448 (2014).
- Chen, L. *et al.* Separation of rare gases and chiral molecules by selective binding in porous organic cages. *Nature Mater.* **13**, 954–960 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was funded by the Leverhulme Trust (F/00 203/T) and by EPSRC (EP/C511794/1). M.G.D.P. acknowledges financial support from ANPCyT (PICT-2011-2128) and from the EC-H2020, MSCRISE-2014 programme, through project 643998 ENACT. L.P. and M.C.G. acknowledge support from the Contrat d'Objectifs Partagés (CNRS, Blaise Pascal University, and the Auvergne Regional Government, France). A.I.C. acknowledges the European Research Council under the European Union's Seventh Framework Programme/ERC Grant Agreement no. 321156 for financial support. We thank M. E. Briggs for assistance with the cage syntheses.

**Author Contributions** N.G. and R.L.G. synthesized the porous crown cage. M.D.P. and G.M. carried out the molecular simulations. K.R. and T.K. performed the PALS measurements. M.C.G. and L.P. measured the methane gas solubilities for the crown cage porous liquid. R.L.G. and A.I.C. conceived the synthesis of the scrambled porous imine cages. R.L.G. synthesized and characterized the scrambled cage porous liquid and measured its gas solubilities. S.L.J. led the project overall and conceived the design of the porous liquid based on the crown-ether cage together with N.G. S.L.J. and A.I.C. led the writing of the manuscript with contributions from all co-authors.

**Additional Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.L.J. (S.James@qub.ac.uk).



## METHODS

**Synthesis of the porous crown-ether cage.** The cage was prepared in six steps. The details for the five-step synthesis of the crown-ether diamine, (15S,16S)-1,4,7,10,13-pentaoxacycloheptadecane-15,16-diamine, are given in Supplementary Information. Reaction of this diamine (8 equiv.) with 1,3,5-triformylbenzene (4 equiv.) in chloroform at 60 °C gave the desired imine cage in a one-pot [4 + 6] cycloimination reaction (Fig. 1a)<sup>6,7</sup>. Melting point, decomposes > 180 °C; IR ( $\nu_{\max}$  in  $\text{cm}^{-1}$ ) 2,858, 1,649, 1,598, 1,445, 1,352, 1,298, 1,250, 1,111, 978, 943, 882;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta_{\text{H}}$  (in p.p.m.) 8.12 (12 H, s), 7.92 (12 H, s), 4.03 (12 H, d,  $J = 7.0$  Hz), 3.75–3.58 (120 H, m);  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta_{\text{C}}$  (in p.p.m.) 161.06, 136.56, 130.20, 73.25, 72.68, 71.70, 71.09, 70.87, 70.60; HRMS (ES+) calc. for  $\text{C}_{108}\text{H}_{156}\text{N}_{12}\text{Na}_2\text{O}_{30}$  [ $\text{M} + 2\text{Na}$ ]<sup>2+</sup> 1,073.5422, found 1,073.5496; CHN analysis calc. for  $\text{C}_{108}\text{H}_{156}\text{N}_{12}\text{O}_{30}$ : C, 61.70; H, 7.48; N, 7.99; found: C, 59.57; H, 7.34; N, 7.74.

**Synthesis of scrambled porous cages.** The scrambled cage mixture was synthesized in one step via the reaction of 1,3,5-triformylbenzene (4 equiv.), 1,2-diamino-2-methylpropane (ref. 23) (13, 3 equiv.) and (*R,R*)-1,2-diaminocyclohexane (ref. 6) (3, 3 equiv.) in dichloromethane at room temperature (yield, 77%). Full details are given in Supplementary Information.

**Preparation of the porous liquids.** The crown cage (0.100 g, 0.0475 mmol, 1 equiv.) was dissolved in 15-crown-5 (113  $\mu\text{l}$ , 0.5707 mmol, 12 equiv.) by sonicating the mixture for 2 h, resulting in a viscous, pale yellow liquid. For the scrambled cages, the solid material was first degassed and then dissolved in rigorously purified hexachloropropene (PCP) to form a pale yellow 10 wt% solution.

**Molecular dynamics (MD) simulations.** Atomistic models for the crown cage were constructed using the OPLS all atom force-field<sup>25</sup>. We computed the free energy cost for introducing a single 15-crown-5 molecule into a cage cavity using the umbrella sampling (US) method<sup>26</sup> combined with the weighted histogram analysis method (WHAM)<sup>27</sup>. The US simulations were performed with the GROMACS-4.5.3 code<sup>28</sup>, in the canonical ensemble (NVT) at 400 K, using a timestep of 0.001 ps. Bulk MD simulations were run on a sample consisting of 40 cages and 480 solvent molecules enclosed in a cubic and periodic simulation box in the isobaric-isothermal ensemble (NPT), using a timestep of 1 fs, a Nosé–Hoover thermostat<sup>29,30</sup>, and a Parrinello–Rahman barostat<sup>31</sup>, with time constants of 0.1 ps and 1 ps, respectively. Lennard–Jones interactions were cut off at 1.4 nm, while Coulomb forces were computed using the smooth particle

mesh Ewald technique<sup>32</sup>. Full computational details are given in Supplementary Information.

**Positron annihilation lifetime spectroscopy (PALS).** Measurements were performed as a function of temperature in a standard experimental set-up, as described previously<sup>33</sup>.

**Gas solubility measurements.** An isochoric technique<sup>34</sup> was used to measure methane gas solubility in the pure 15-crown-5 solvent and in the porous liquid. For the scrambled cages, gas solubilities were either measured by  $^1\text{H}$  NMR (for example, for methane) or, for larger bulk samples of the porous liquid, by displacing the gas by addition of one molar equivalent (relative to the cage) of a second, small guest (for example, chloroform, Fig. 4b) and measuring the amount of gas (for example, xenon) evolved.

25. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225 (1996).
26. Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **91**, 275 (1995).
27. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **13**, 1011 (1992).
28. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435 (2008).
29. Nosé, S. A molecular-dynamics methods for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255 (1984).
30. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695 (1985).
31. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182 (1981).
32. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577 (1995).
33. Harms, S. *et al.* Free volume of interphases in model nanocomposites studied by positron annihilation lifetime spectroscopy. *Macromolecules* **43**, 10505 (2010).
34. Jacquemin, J., Costa Gomes, M. F., Husson, P. & Majer, V. Solubility of carbon dioxide, ethane, methane, oxygen, nitrogen, hydrogen, argon, and carbon monoxide in 1-butyl-3-methylimidazolium tetrafluoroborate between temperatures 283 K and 343 K and at pressures close to atmospheric. *J. Chem. Thermodyn.* **38**, 490 (2006).

# Plate tectonics on the Earth triggered by plume-induced subduction initiation

T. V. Gerya<sup>1</sup>, R. J. Stern<sup>2</sup>, M. Baes<sup>3</sup>, S. V. Sobolev<sup>3,4</sup> & S. A. Whattam<sup>5</sup>

Scientific theories of how subduction and plate tectonics began on Earth—and what the tectonic structure of Earth was before this—remain enigmatic and contentious<sup>1</sup>. Understanding viable scenarios for the onset of subduction and plate tectonics<sup>2,3</sup> is hampered by the fact that subduction initiation processes must have been markedly different before the onset of global plate tectonics because most present-day subduction initiation mechanisms require acting plate forces and existing zones of lithospheric weakness, which are both consequences of plate tectonics<sup>4</sup>. However, plume-induced subduction initiation<sup>5–9</sup> could have started the first subduction zone without the help of plate tectonics. Here, we test this mechanism using high-resolution three-dimensional numerical thermomechanical modelling. We demonstrate that three key physical factors combine to trigger self-sustained subduction: (1) a strong, negatively buoyant oceanic lithosphere; (2) focused magmatic weakening and thinning of lithosphere above the plume; and (3) lubrication of the slab interface by hydrated crust. We also show that plume-induced subduction could only have been feasible in the hotter early Earth for old oceanic plates. In contrast, younger plates favoured episodic lithospheric drips rather than self-sustained subduction and global plate tectonics.

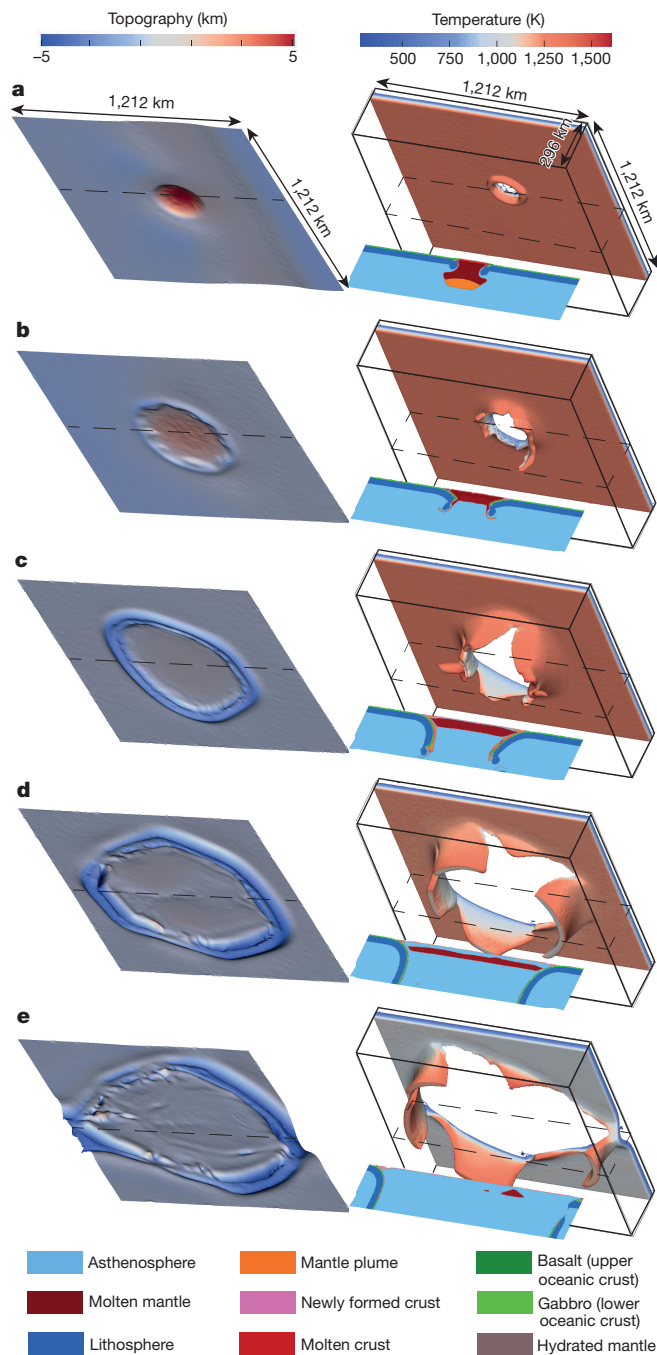
Plate tectonics involves the independent movement of lithospheric plates, which is mostly powered by the sinking of dense lithosphere in subduction zones; therefore, understanding how plate tectonics began requires an understanding of how the first subduction zone formed. This in turn requires resolving the oceanic lithospheric buoyancy-strength paradox: lithosphere becomes both denser (favouring sinking) and stronger (impeding rupture) as it ages, cools and thickens. Gravitational spreading of continents at passive margins has been proposed to be a kick-starter of global plate tectonics in the hotter early Earth<sup>3</sup>, but it remains debatable<sup>8,10</sup> how early continental crust formed and whether large, gravitationally unstable continents with thick crust could have formed without plate tectonics. It is also unclear whether continents are prerequisites for subduction initiation; most Cenozoic-era subduction zones initiated within oceanic lithosphere<sup>11</sup>. Another paradox is that lithospheric weaknesses are needed to initiate subduction, but today such weaknesses are produced by ongoing plate tectonics. Exogenic processes such as meteorite impacts have been suggested<sup>12</sup> as mechanisms for subduction initiation; however, impacts sufficiently energetic to rupture the lithosphere probably only occurred very early in the history of Earth. Self-organization of weak proto-plate boundaries by ductile damage of stagnant lithosphere in zones of convective downwellings has been proposed<sup>2</sup> as a way to start plate tectonics. This hypothesis, however, does not address subduction initiation itself, which is required for plate tectonics to begin. Recognition that the Caribbean Large Igneous Province (CLIP), formed by the arrival of a large plume head 100–95 million years (Myr) ago, might have induced new subduction zones<sup>7</sup> encouraged us to explore whether a similar mechanism might have been responsible for initiating subduction<sup>13</sup> and plate tectonics<sup>8</sup> during the Precambrian eon (>542 Myr ago).

Plume-induced subduction initiation has been previously investigated by means of two-dimensional (2D) thermomechanical modelling<sup>6</sup>. However, plume–lithosphere interaction is an intrinsically three-dimensional (3D) process and it remains unclear whether plume-induced subduction initiation is feasible for finger-like and mushroom-like purely thermal plumes. We expect that the most critical 3D effect is ring confinement of the circular slab, which forms around the plume head; unless the ring ruptures, it precludes further sinking of the slab into the mantle. It also remains uncertain how plume–lithosphere interaction and related subduction initiation would have differed in the Precambrian Earth, where hotter upper mantle, and consequently a higher degree of melting, modified oceanic lithosphere properties, especially crustal thickness and upper-mantle rheology and density<sup>1</sup>. Here, we designed a 3D high-resolution plume–lithosphere interaction model (Extended Data Fig. 1, Extended Data Table 1) that allows for self-organization of new plate boundaries (Methods). The model takes into account melt extraction and upward transport from the plume, rheological weakening of the lithosphere subjected to melt percolation<sup>14</sup> and crustal growth by magmatic processes (Methods).

Typical model evolution leading to self-sustaining subduction (Extended Data Table 2) is subdivided into five stages (Fig. 1): (1) oceanic plateau formation by arrival of a mantle plume head; (2) formation of an incipient trench and a descending nearly circular slab at the plateau margins; (3) tearing of the circular slab under its own weight; (4) formation of several self-sustained retreating subduction zones; and (5) cooling and spreading of the new lithosphere formed between the retreating subduction zones. At the first stage (Fig. 1a), oceanic plateau formation by magmatic processes associates with rheological weakening and thinning of the underlying lithosphere, thus allowing for plume propagation towards the surface. This plume propagation enhances magma production by decompression melting and accelerates crustal growth atop the plume, leading to rapid plateau thickening and uplift. Plume penetration and plateau growth induce substantial lateral density contrasts at the plateau margins. This contrast in density creates favourable conditions for subduction initiation by the same type of gravitational instability that triggers subduction along oceanic transform boundaries<sup>4,11</sup>. The instability reflects the juxtaposition of old, cold, negatively buoyant pre-existing oceanic lithosphere and young, warm, positively buoyant thinned lithosphere of the oceanic plateau. At the second stage (Fig. 1b), the oceanic plateau begins to collapse and thrusts radially outward over the surrounding colder oceanic plate, which gradually bends down and penetrates into the mantle, thus forming an incipient nearly circular lithospheric slab. Slab descent is increasingly resisted by the ring confinement at the base of the encircling slab. At the third stage (Fig. 1c), the ring confinement is overcome by tearing the slab. The tears become self-sustained, owing to stress focusing at the tips of the propagating tears combined with the non-Newtonian strain-weakened viscous-plastic mantle rheology (Methods). Slab tearing introduces asymmetry into the initially nearly circular subduction system, splitting it into several independently retreating segments with

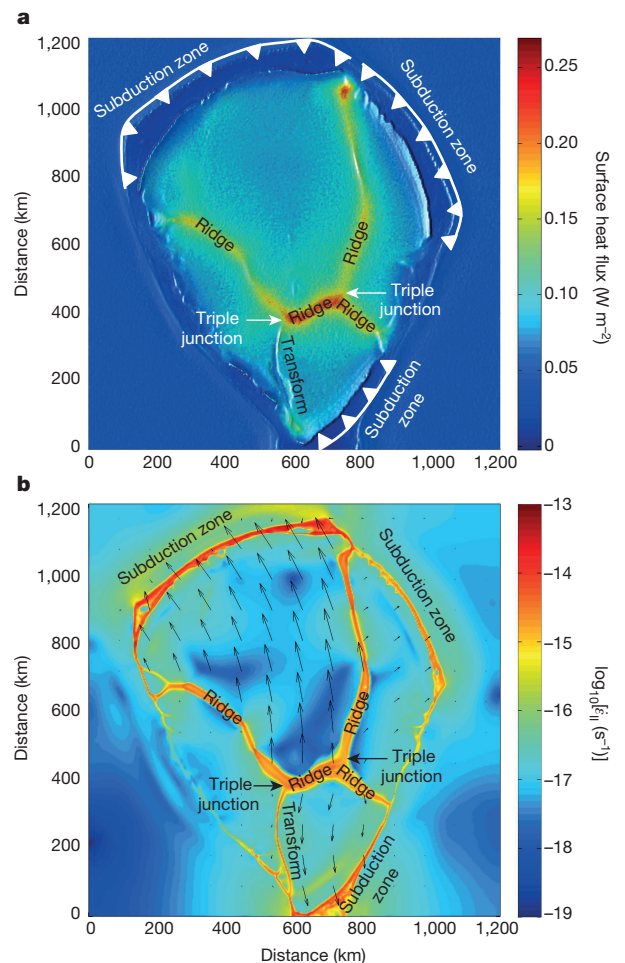
<sup>1</sup>Department of Earth Sciences, ETH-Zurich, Sonneggstrasse 5, 8092 Zurich, Switzerland. <sup>2</sup>Geosciences Department, University of Texas at Dallas, Richardson, Texas 75083-0688, USA.

<sup>3</sup>GFZ German Research Center for Geosciences, Heinrich-Mann-Allee 18/19, 14473 Potsdam, Germany. <sup>4</sup>Institute of Earth and Environmental Science, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam-Golm, Germany. <sup>5</sup>Department of Earth and Environmental Sciences, Korea University, Seoul 136-701, South Korea.



**Figure 1 | Dynamics of plume-induced subduction initiation under present-day mantle temperature conditions.** The model used for the simulations was 'bsayea'; see Extended Data Table 2. **a**, Oceanic plateau development (at 0.07 Myr from the beginning of the experiment); **b**, formation of an incipient trench and a nearly circular slab (at 0.56 Myr); **c**, tearing of the circular slab (at 3.79 Myr); **d**, formation of retreating subduction zones (at 9.43 Myr); and **e**, development of spreading centres and transform boundaries (at 25.63 Myr). Left and right columns show topography and subducted lithosphere morphology with projected slab surface temperature, respectively. Dashed lines indicate positions of the 2D cross-sections (colour code is at the bottom of the figure).

different orientations and dip angles. At the fourth stage (Fig. 1d), subduction becomes self-sustained and retreats towards the boundaries of the model. The initially elevated, thick oceanic plateau collapses as it extends within the still-widening and cooling region of new oceanic lithosphere formed between retreating subduction zones. Finally, at the fifth stage (Fig. 1e), a mosaic of nearly independently moving, growing

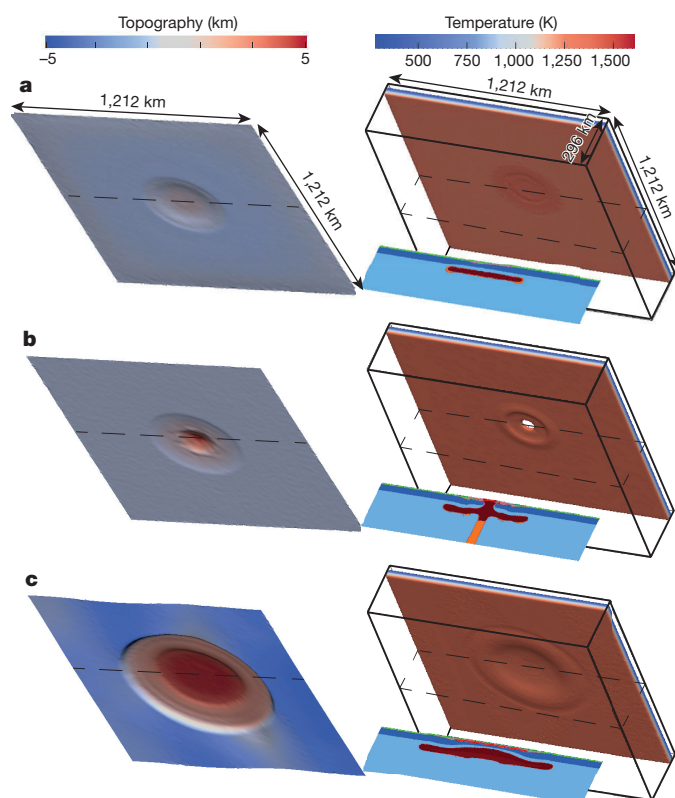


**Figure 2 | Development of an embryonic mosaic of plates separated by spreading centres (ridges), triple junctions and transform faults at the latest stage of plume-induced subduction.** The model used for the simulations was 'bsayea' at 25.63 Myr; see Fig. 1e and Extended Data Table 2. **a**, Surface heat fluxes projected onto the modelled surface topography, showing a pattern of spreading centres (white lines with triangles show dip directions of retreating subducting slabs). **b**, Spatial distribution of the second strain rate invariant ( $\dot{\epsilon}_{II}$ ), where the overdot indicates differentiation with respect to time at a depth of 20 km (arrows show horizontal velocities of individual, young, non-subducting plates moving towards retreating subducting slabs).

and cooling small, young, non-subducting plates heading towards individual retreating subduction zones forms in the widening oceanic region (Fig. 2). The plates are separated by spreading centres (ridges), triple junctions and transform faults, and thus the newly formed multi-slab subduction system operates as an embryonic plate-tectonic cell.

The critical parameter affecting the feasibility of plume-induced subduction initiation is the magnitude of magmatic weakening  $\lambda_{\text{melt}}$  (Methods) of lithosphere above the plume, which controls the ability of the plume to penetrate the lithosphere and thus create the initial gravitational instability. When melt-induced weakening is insufficient ( $\lambda_{\text{melt}} \gg 0.01$ ), the plume head cannot penetrate the lithosphere, but instead spreads at the lithosphere–asthenosphere boundary, and upwelling stops. The plume head produces thickened crust and an elevated oceanic plateau, but does not trigger subduction (Fig. 3a). In this situation, a long-lived plume supplying hot mantle material to the bottom of the lithosphere favours lithospheric thinning even at the reduced intensity of magmatic weakening (Fig. 3b), thus creating favourable conditions for subduction initiation. A similar effect is obtained by increasing the size and temperature of the plume head (Fig. 3c), which increases its buoyancy and thus induces larger stresses in the overlying

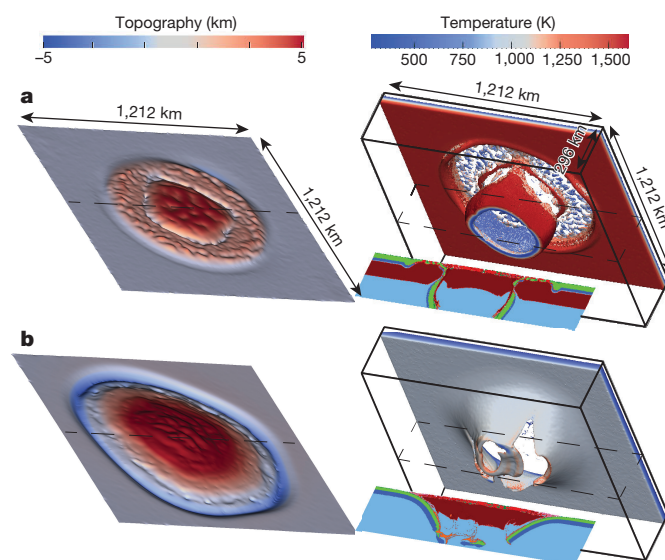




**Figure 3 | Influence of model parameters on the thinning of the oceanic plate under a condition of reduced intensity of magmatism-induced lithospheric weakening.** Compare with the reference model shown in Fig. 1. **a**, Failed subduction initiation caused by insufficient magmatism-induced weakening of the plate (model, 'bsasb' at 7.32 Myr). **b**, Influence of a long-lived plume with a conduit providing additional heat and mass supply through time (model, 'bsan' at 2.00 Myr). **c**, Influence of increased plume size and temperature (model, 'bsba' at 1.24 Myr). Colour code is the same as that used in Fig. 1. See Extended Data Table 2 for details of the models.

lithosphere<sup>6</sup>. The age of the oceanic plate has a more complex influence on subduction initiation (Extended Data Fig. 2). On the one hand, very young (<20 Myr old) lithosphere does not favour self-sustained subduction because the negative buoyancy of such plates is insufficient to provide the necessary slab pulling force<sup>1</sup> (Extended Data Fig. 2a). On the other hand, plates that are very old (>80 Myr old), and thus cold and strong, are more difficult for a plume head to penetrate and initiate subduction (Extended Data Fig. 2d). Therefore, older lithosphere needs larger and hotter plumes for subduction initiation. Sustainability of subduction also requires rheologically weak, hydrated upper oceanic crust, which lubricates the incipient subduction interface (Methods). In the absence of such crust, subduction freezes soon after initiation (Extended Data Fig. 3). A similar subduction-freezing effect can also be achieved (Extended Data Table 2) if we assume a stronger, dry rheology of the plume or compositional buoyancy of the pre-existing oceanic lithosphere<sup>1</sup> (Methods), caused by dehydration and depletion by melt extraction<sup>15</sup>. This freezing effect can, however, be overcome by assuming greater plate ages (>30 Myr) or faster hydration (weakening) of the mantle wedge by subduction-released fluids (Extended Data Table 2).

The Archaean mantle was 100–300 K hotter than is the modern mantle<sup>16</sup>, which would have produced depleted compositionally buoyant lithosphere (Methods) with 10–30-km-thick oceanic crust<sup>1</sup>. This crust thickness strongly increases buoyancy and thus decreases the 'subductability' of the oceanic lithosphere<sup>1</sup>. In addition, the crust–mantle boundary ('Moho') becomes deeper and hotter, which decreases the viscosity of rocks at this boundary and leads to rheological



**Figure 4 | Plume–lithosphere interaction for hotter mantle temperature and thicker oceanic crust.** **a**, Development of plume-induced lithospheric drips for 20-Myr-old oceanic plate with 30-km-thick crust (model, 'bsar' at 0.34 Myr; see Extended Data Table 2 and Extended Data Fig. 4). **b**, Development of plume-induced self-sustaining subduction for 80-Myr-old oceanic plate with 20-km-thick crust (model, 'bsayh' at 2.50 Myr; see Extended Data Table 2). Colour code is the same as that used in Fig. 1.

weakening of the plate. Here, we explore two types of Archaean oceanic plates: (1) a young, warm, positively buoyant and weak plate, which is not favourable for subduction<sup>1</sup> (Fig. 4a) and (2) an old, cold, negatively buoyant and strong plate, which is favourable for subduction<sup>1</sup> (Fig. 4b).

Plume interaction with young lithosphere results in oceanic plateau formation followed by nearly circular sheet-like lithospheric 'drips' that are driven by thickened lower-crust eclogitization and are terminated by shallow circular necking (break-off) of the descending oceanic lithosphere (Extended Data Fig. 4). Surface expression of the initial circular drip is similar to that of true subduction initiation with a pronounced trench and collapsing oceanic plateau (compare with Fig. 1b and Extended Data Fig. 4b); however, eclogite dripping is short-lived and does not create coherent retreating slabs. Repetitive drips quickly remove pre-existing mantle lithosphere, so that thickened, hot mafic crust comes to overlie convecting asthenosphere (Extended Data Fig. 4d). An intense exchange of heat and mass then occurs between crust and mantle: mantle-derived melt additions contribute to further growth of the crust as eclogitized lower crust sinks into the mantle and re-melts. In contrast, plume interaction with old lithosphere results in self-sustained subduction, which is further assisted by strong densification of thick oceanic crust, owing to its eclogitization (Fig. 4b). This observation suggests that plume-induced subduction could have occurred for hot Archaean mantle if oceanic lithosphere at that time was sufficiently old, cold and thus strong and negatively buoyant<sup>1</sup>. Compositional buoyancy of the depleted Archaean oceanic lithosphere with thick crust does not preclude subduction (Extended Data Table 2), but requires greater cooling ages (>60–70 Myr)<sup>1</sup> for plate subductability compared to present-day mantle temperature conditions (>10–30 Myr)<sup>1</sup>.

Our results show that the arrival of a sufficiently large, hot and long-lived mantle plume could weaken strong and dense negatively buoyant<sup>1</sup> oceanic lithosphere to produce new self-sustaining subduction zones, which are required for the beginning of plate tectonics. In addition, a global plate mosaic is needed to initiate modern-style plate tectonics with various simultaneously operating plate boundaries<sup>2</sup>. Some possibilities for creating such a plate mosaic arise from plume-induced subduction initiation. The newly formed oceanic lithosphere above the plume head makes up the first rectangular plate<sup>2</sup> (Fig. 1d), or even an embryonic plate mosaic (Fig. 2). Multiple plates can also be

formed by at least three complementary mechanisms. First, they can be created by the interaction of multiple subduction zones produced by several simultaneous mantle plumes during enhanced mantle activity periods, perhaps corresponding to peaks of new mantle-derived crust formation<sup>10</sup>. Second, crustal and lithospheric thickness heterogeneities can result from failed subduction initiation episodes creating multiple elevated oceanic plateaus with thickened crust and thinned mantle lithosphere (Fig. 3), which could later localize plate stresses to form subduction zones. Third, rheologically weak, 'frozen' subduction zones, spreading centres and transform faults can be preserved in the lithosphere from past attempted episodes of subduction (Figs 1 and 2) and/or lithospheric drip (Fig. 4a) if these episodes terminated at some intermediate stage (Extended Data Fig. 3) and did not result in global lithospheric lid resurfacing. Owing to incomplete healing of the deformation-related damage, these boundaries can be reactivated after subduction re-initiation<sup>2</sup>. These boundaries would be the first plate-tectonic lithospheric weaknesses that could be exploited, allowing formation of more subduction zones, spreading centres and transform faults, progressively infecting the entire lid.

Several questions remain to be addressed before this hypothesis can be accepted. The paramount question is why the arrival of other oceanic plume heads that formed large igneous provinces (LIPs) have not caused new subduction zones to form around them. The Caribbean is the only one of the eight Mesozoic and Cenozoic oceanic plateaus<sup>17</sup> that is thought to have triggered subduction around its margins<sup>7</sup>. Although collapse of surrounding oceanic lithosphere requires a large plume head, the size of the plume head does not seem to be the only controlling factor for whether or not a plume will trigger a flanking subduction zone. Three other aspects of the CLIP favour subduction initiation: (1) old age (>40 Myr) and hence large negative buoyancy<sup>1</sup> of the precursor oceanic lithosphere; (2) longevity of plume–lithosphere interaction (duration of at least 56 Myr)<sup>18</sup>; and (3) very high plume temperature (up to 1,620 °C for associated Gorgona komatiites)<sup>19</sup>. Both the longevity and the plume temperature are maximal among Mesozoic oceanic hot spots and LIPs and thus caused larger melt flux through the lithosphere, which favoured its strong weakening (Methods). Other prominent oceanic LIPs, such as the Ontong Java Plateau and Shatsky Rise were formed on younger (<20 Myr old) and warmer oceanic lithosphere, compared to the CLIP. According to our 3D experiments, such warm lithosphere either should not subduct at all (Extended Data Fig. 2a) or should generate lithospheric drips without self-sustained subduction (Fig. 4a).

Our numerical experiments confirm that only the most energetic mantle plumes with large temperature excess feeding a large plume head are able to weaken the lithosphere enough to initiate subduction. Such plumes were probably present in the hotter Archaean mantle, and probably produced thick magmatic plateaus and possibly triggered the onset of global modern-style plate tectonics<sup>8</sup>. Although the thermal evolution of the Earth implies that plume activities in the past might have been less intense than at present<sup>20</sup>, geological records suggest a substantial level of plume activity during the Precambrian eon<sup>8,21–23</sup>. It has yet to be analysed what evidence in the rock record supports the association of subduction initiation with plumes in the Precambrian. In contrast to Venus, where plume-induced subduction-like features are observed for large circular surface structures (coronae)<sup>5,6,8</sup>, Earth does not offer an unperturbed record of Precambrian crust. Therefore, geological and geochemical data must be used instead<sup>8–10,13,24,25</sup>. Dhuime *et al.*<sup>10</sup> studied systematic variations in hafnium and oxygen isotopes in zircons of different ages that reveal the relative proportions of reworked crust and of new mantle-derived (juvenile) crust through time. They showed that several Precambrian episodes of very strong mantle activity correspond to peaks in juvenile crust production, the largest of which preceded a hypothesized Archaean transition to plate tectonics<sup>10</sup>. Geodynamical processes that might have dominated crustal growth before plate tectonics include a wide range of Venus-like active stagnant-lid behaviour, including mantle plumes, lithospheric

delamination and drips, or 'intraplate' lithospheric extension or mantle upwelling<sup>8,10</sup>; these processes are consistent with our numerical results for hot mantle temperatures and young lithospheric plates (Fig. 4a). On the basis of a synthesis of geological–geochemical data, it was proposed in refs 8 and 9 that the formation of large plume-derived plateaus in the Archaean eon might have led to the onset of subduction and plate tectonics. The proposed physical mechanism is similar to plume-induced subduction initiation and implies gravitational collapse of the growing plateaus, leading to lithospheric subduction along their margins. As in our numerical models, the onset of subduction would then have generated arc-like magmas beneath the margins of the plateaus<sup>9,26</sup> and other plate-tectonic structures, such as spreading centres and transform faults in their extending interiors<sup>8</sup>. Geochemical evidence from Precambrian greenstone belts suggests close spatial and temporal associations between arc-like and plume-like signatures<sup>13,24,25</sup>. This evidence in particular suggests that short-lived (5–10 Myr) discrete subduction-like events were superimposed on background plume-like mantle activity<sup>13</sup>. Whereas arc-like signatures are not exclusive to subduction zones<sup>27,28</sup>, and alternative interpretations are proposed for arc-like igneous rocks in some Archaean greenstone belts<sup>29,30</sup>, apparent short-term evolution from plume-dominated to arc-like crustal growth<sup>13</sup> provides geological evidence for a link between plume activity and subduction initiation in the Precambrian. Depending on lithospheric strength, such plume–arc interactions could reflect either plume-induced self-sustained subduction (Fig. 4b) or plume-induced episodic drips (Fig. 4a), which are expected for younger lithosphere. In either case, these events were short-lived<sup>13</sup>, leading to short subduction-like episodes. The geodynamic transition from global plume-lid tectonics with plume-induced short-lived lithospheric drips and embryonic subduction zones to global modern-style plate tectonics reflect thickening, strengthening and densification of the lithosphere as the result of secular cooling of the upper mantle. In addition to increasing strength and negative buoyancy<sup>1</sup> of oceanic lithosphere with decreasing upper mantle temperature, which promoted long-lasting self-sustained subduction<sup>13</sup>, accumulation of lithospheric heterogeneities and weakness zones<sup>2</sup> and plume-assisted continental growth<sup>3,8</sup> might have also influenced when and how the transition from short-lived Archaean plume–subduction interactions to modern-style plate tectonics occurred.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 13 February; accepted 15 September 2015.**

- Korenaga, J. Initiation and evolution of plate tectonics on Earth: theories and observations. *Annu. Rev. Earth Planet. Sci.* **41**, 117–151 (2013).
- Bercowski, D. & Ricard, Y. Plate tectonics, damage and inheritance. *Nature* **508**, 513–516 (2014).
- Rey, P. F., Coltice, N. & Flament, N. Spreading continents kick-started plate tectonics. *Nature* **513**, 405–408 (2014).
- Stern, R. J. Subduction initiation: spontaneous and induced. *Earth Planet. Sci. Lett.* **226**, 275–292 (2004).
- Sandwell, D. T. & Schubert, G. Evidence for retrograde lithospheric subduction of Venus. *Science* **257**, 766–770 (1992).
- Ueda, K., Gerya, T. & Sobolev, S. V. Subduction initiation by thermal–chemical plumes: numerical studies. *Phys. Earth Planet. Inter.* **171**, 296–312 (2008).
- Whattam, S. A. & Stern, R. J. Late Cretaceous plume-induced subduction initiation along the southern margin of the Caribbean and NW South America: the first documented example with implications for the onset of plate tectonics. *Gondwana Res.* **27**, 38–63 (2015).
- Van Kranendonk, M. J. Two types of Archaean continental crust: plume and plate tectonics on early Earth. *Am. J. Sci.* **310**, 1187–1209 (2010).
- Nair, R. & Chacko, T. Role of oceanic plateaus in the initiation of subduction and origin of continental crust. *Geology* **36**, 583–586 (2008).
- Dhuime, B., Hawkesworth, C. J., Cawood, P. A. & Storey, C. D. A change in the geodynamics of continental growth 3 billion years ago. *Science* **335**, 1334–1336 (2012).
- Gurnis, M., Hall, C. & Lavier, L. Evolving force balance during incipient subduction. *Geochim. Geophys. Geosyst.* **5**, Q07001 (2004).
- Hansen, V. L. Subduction origin on early Earth: a hypothesis. *Geology* **35**, 1059–1062 (2007).
- Moyen, J.-F. & van Hunen, J. Short-term episodicity of Archaean plate tectonics. *Geology* **40**, 451–454 (2012).

14. Keller, T., May, D. A. & Kaus, B. J. P. Numerical modelling of magma dynamics coupled to tectonic deformation of lithosphere and crust. *Geophys. J. Int.* **195**, 1406–1442 (2013).
15. Morgan, J. P., Morgan, W. J. & Price, E. Hotspot melting generates both hotspot volcanism and a hotspot swell? *J. Geophys. Res.* **100**, 8045–8062 (1995).
16. Herzberg, C., Condie, K. & Korenaga, J. Thermal history of the Earth and its petrological expression. *Earth Planet. Sci. Lett.* **292**, 79–88 (2010).
17. Eldholm, O. & Coffin, M. F. in *The History and Dynamics of Global Plate Motion* (eds Richards, M. A. *et al.*) 309–326 (Geophysical Monograph Series, Vol. 121, AGU, 2000).
18. Hoernle, K., Hauff, F. & van den Bogaard, P. 70 m.y. history (139–69 Ma) for the Caribbean large igneous province. *Geology* **32**, 697–700 (2004).
19. Herzberg, C. & Gazel, E. Petrological evidence for secular cooling in mantle plumes. *Nature* **458**, 619–622 (2009).
20. Jellinek, A. M. & Manga, M. Links between long-lived hot spots, mantle plumes, D", and plate tectonics. *Rev. Geophys.* **42**, RG3002 (2004).
21. Condie, K. C. & Benn, K. in *Archean Geodynamics and Environments* (eds Benn, K. *et al.*) 47–59 (Geophysical Monograph Series, Vol. 164, AGU, 2006).
22. Campbell, I. H. in *Mantle Plumes: Their Identification Through Time* (eds Ernst, R. E. & Buchan, K. L.) 5–21 (Geological Society of America Special Paper 352, GSA, 2001).
23. Olsson, J. R., Söderlund, U., Hamilton, M. A., Klausen, M. B. & Helffrich, G. R. A late Archaean radiating dyke swarm as possible clue to the origin of the Bushveld Complex. *Nature Geosci.* **4**, 865–869 (2011).
24. Puchtel, I. S. *et al.* Combined mantle plume-island arc model for the formation of the 2.9 Ga Sumo-Keonozo greenstone belt, SE Baltic Shield: isotope and trace element constraints. *Geochim. Cosmochim. Acta* **63**, 3579–3595 (1999).
25. Wyman, D. A., Kerrich, R. & Polat, A. Assembly of Archean cratonic mantle lithosphere and crust: plume–arc interaction in the Abitibi–Wawa subduction–accretion complex. *Precamb. Res.* **115**, 37–62 (2002).
26. Smithies, R. H., Champion, D. C., Van Kranendonk, M. J., Howarda, H. M. & Hickmana, A. H. Modern-style subduction processes in the Mesoarchaean: geochemical evidence from the 3.12 Ga Whundo intra-oceanic arc. *Earth Planet. Sci. Lett.* **231**, 221–237 (2005).
27. Morris, G. A., Larson, P. B. & Hooper, P. R. 'Subduction style' magmatism in a non-subducting setting: the Colville Igneous Complex, NE Washington State, USA. *J. Petrol.* **41**, 43–67 (2000).
28. Willbold, M., Hegner, E., Stracke, A. & Rocholl, A. Continental geochemical signatures in dacites from Iceland and implications for models of early Archaean crust formation. *Earth Planet. Sci. Lett.* **279**, 44–52 (2009).
29. Barnes, S. J. & Van Kranendonk, M. J. Archean andesites in the east Yilgarn craton, Australia: products of plume–crust interaction? *Lithosphere* **6**, 80–92 (2014).
30. Champion, D. C. & Smithies, R. H. in *Earth's Oldest Rocks* (eds van Kranendonk, M. J. *et al.*) 369–409 (Developments in Precambrian Geology, Vol. 15, Elsevier, 2007).

**Acknowledgements** This study was co-funded by the ERC ITN project ZIP (T.V.G.), the SNF project Swiss-AlpArray (T.V.G.), the SNF grant 200021\_149252 (T.V.G.), the ETH grant ETH-37\_11-2 (T.V.G.) and a SNF short scientific visits program (R.J.S.). Simulations were performed on the ETH-Zurich Brutus cluster and on the GFZ-Potsdam cluster. Open-source software ParaView (<http://www.paraview.org>) was used for 3D visualization.

**Author Contributions** T.V.G. designed the study, conducted part of the numerical experiments, interpreted the results and designed the 3D thermo-mechanical code. R.J.S. designed the study, analysed the natural data and interpreted the results. M.B. conducted part of the numerical experiments and interpreted the results. S.V.S. designed the study and interpreted the results. S.A.W. analysed the natural data and interpreted the results. All authors discussed the results, problems and methods, interpreted the data and wrote the paper.

**Additional Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.V.G. ([taras.gerya@erdw.ethz.ch](mailto:taras.gerya@erdw.ethz.ch)).



## METHODS

**Modelling of subduction initiation.** *Modelling approach.* The numerical thermo-mechanical 3D C code I3ELVIS used for the modelling of subduction initiation is based on a combination of a finite-difference method, applied on a staggered Eulerian grid, and a marker-in-cell technique<sup>31,32</sup>. The momentum, mass and energy conservation equations are solved on the Eulerian frame, and physical properties are transported by Lagrangian markers that move according to the velocity field interpolated from the fixed grid. Non-Newtonian viscous-plastic rheologies are used in the model (Extended Data Table 1), which is also fully thermodynamically coupled and accounts for melting processes as well as adiabatic, radiogenic, latent and frictional internal heating sources. Full details of this method, allowing for its reproduction, are provided elsewhere<sup>31,32</sup>.

*Numerical model design.* The initial model set-up places a thermal mantle plume (100–300 km in diameter) below an oceanic lithosphere with 8–30-km-thick mafic crust composed of 3-km basalts and 5–27-km gabbro (Extended Data Fig. 1, Extended Data Table 1). The Eulerian computational domain is equivalent to 1,212 km × 1,212 km × 296 km (Extended Data Fig. 1) and is resolved with a regular rectangular grid of 405 × 405 × 149 nodes and contains 200 million randomly distributed Lagrangian markers. The model imposes free-slip mechanical boundary conditions on all except the lower boundary, which is open for mantle flow. The free-surface boundary condition atop the crust is implemented by using a 20-km-thick ‘sticky’ air/water layer<sup>33</sup> with low density (1 kg m<sup>-3</sup> at 0–16 km from the top of the model, 1,000 kg m<sup>-3</sup> below 16 km) and reduced<sup>33</sup> viscosity (10<sup>18</sup> Pa s). The initial thermal structure and thickness of the plate (Extended Data Fig. 1) is defined by prescribing its laterally uniform cooling age<sup>1,34</sup> and respective geotherm<sup>34</sup> with 273 K at the surface and the mantle potential temperature of 1,573–1,773 K varied in different experiments (Extended Data Table 2, Extended Data Figs 2–4). An adiabatic thermal gradient of 0.5 K km<sup>-1</sup> is initially prescribed in the asthenospheric mantle. The simplified, purely thermal, spherical mantle plume is prescribed in the centre of the model with elevated (by 100–300 K) temperature compared to the surrounding mantle (Extended Data Fig. 1). Temperature-dependent thermal conductivity is used for the mantle and the crust (Extended Data Table 1). The thermal boundary conditions are 273 K at the top, 1,711–1,911 K (depending on the mantle potential temperature) at the bottom and zero heat flux on all other boundaries. To ensure an efficient heat transfer from the surface of the crust, the temperature of the ‘sticky’ air/water is kept constant at 273 K. Gravitational acceleration of 9.81 m s<sup>-2</sup> was used in the model.

*Viscous-plastic rheological model.* The viscous and brittle (plastic) properties (Extended Data Table 1) are implemented via evaluation of the effective viscosity of the material. For the ductile materials, the contributions from different flow laws such as dislocation and diffusion creep are taken into account via computation of the inverse average ductile viscosity  $\eta_{\text{ductile}}$

$$\frac{1}{\eta_{\text{ductile}}} = \frac{1}{\eta_{\text{newt}}} + \frac{1}{\eta_{\text{powl}}}$$

$\eta_{\text{newt}}$  and  $\eta_{\text{powl}}$  are the effective viscosities for diffusion and dislocation creep, respectively, computed as

$$\eta_{\text{newt}} = \frac{A_D}{2\sigma_{\text{cr}}^{n-1}} \exp\left(\frac{E + PV}{RT}\right)$$

$$\eta_{\text{powl}} = \frac{A_D^{1/n}}{2} \exp\left(\frac{E + PV}{nRT}\right) \dot{\epsilon}_{\text{II}}^{1/n-1}$$

where  $P$  is pressure,  $T$  is temperature (in kelvin),  $\dot{\epsilon}_{\text{II}} = \sqrt{\dot{\epsilon}_{ij}\dot{\epsilon}_{ij}/2}$  is the second invariant of the strain rate tensor (where the overdot indicates differentiation with respect to time and the repeated indices  $i, j$  here and elsewhere imply summation:  $\dot{\epsilon}_{ij}\dot{\epsilon}_{ij} = \sum_{i=x,y,z} \sum_{j=x,y,z} \dot{\epsilon}_{ij}\dot{\epsilon}_{ij}$ ),  $\sigma_{\text{cr}}$  is the diffusion-dislocation transition stress<sup>34</sup>, and  $A_D$ ,  $E$ ,  $V$  and  $n$  are experimentally determined flow-law parameters (Extended Data Table 1), corresponding to the material constant, the activation energy, the activation volume and the stress exponent, respectively. Relatively weak, wet mantle rheology was used for the plume; however, models with stronger, dry mantle rheology were also tested (Extended Data Table 2) to account for plume dehydration due to melt extraction<sup>15</sup>.

The ductile rheology is combined with a brittle (plastic) rheology to yield an effective viscous-plastic rheology using the following upper limit for the ductile viscosity

$$\eta_{\text{ductile}} \leq \frac{C + \phi P \lambda_{\text{melt}}}{2\dot{\epsilon}_{\text{II}}}, \quad (1)$$

$$\phi = \begin{cases} \phi_0(1 - \gamma/\gamma_0) & \text{for } \gamma \leq \gamma_0 \\ 0 & \text{for } \gamma > \gamma_0 \end{cases},$$

$$\gamma = \int \sqrt{\dot{\epsilon}_{ij}(\text{plastic})\dot{\epsilon}_{ij}(\text{plastic})/2} dt$$

where  $\lambda_{\text{melt}} = 1 - P_{\text{melt}}/P$  is the assumed long-term melt-induced weakening factor<sup>35</sup>, which depends on the ratio between the melt pressure ( $P_{\text{melt}}$ ) and total pressure ( $P$ )<sup>14</sup>,  $\phi_0$  is the internal friction coefficient ( $\phi_0$  is the initial internal friction coefficient; Extended Data Table 1),  $\gamma \geq 0$  is the time-integrated plastic strain ( $\gamma_0 = 0.5$  is the upper strain limit for the fracture-related weakening)<sup>36</sup>,  $C$  is the tensile rock strength at  $P = 0$  (Extended Data Table 1),  $t$  is time (in seconds) and  $\dot{\epsilon}_{ij}(\text{plastic})$  is the plastic strain rate tensor. We stress that melt-induced weakening ( $\lambda_{\text{melt}} < 1$ , Extended Data Table 2) is applied locally on lithospheric rock markers above areas of melt extraction computed at a given time step. For all other rocks, no magmatic weakening has been assumed ( $\lambda_{\text{melt}} = 1$ ). We also point out that our large-scale viscous-plastic rheological model of the lithosphere is simplified and does not account for rock elasticity and ductile damage processes<sup>2</sup>, nor does it account for healing of deactivated fractures<sup>2,36</sup>.

*Partial melting and melt extraction model.* Melting of the mantle plume and the asthenosphere, as well as melt extraction and percolation across the crust–mantle interface and to the surface is implemented in a simplified manner<sup>36</sup>. According to our model, mafic magma added to the crust is balanced by melt production and extraction in the plume. However, melt percolation<sup>37</sup> is not modelled directly and is considered to be nearly instantaneous<sup>38</sup>. The standard (that is, without melt extraction) volumetric degree of mantle melting,  $M_0$ , changes with pressure and temperature according to the parameterized batch melting model of ref. 39. Lagrangian markers track the amount of melt extracted during the evolution of each experiment. The total amount of melt,  $M$ , for every marker takes into account the amount of previously extracted melt and is calculated as<sup>40</sup>  $M = M_0 - \sum_m M_{\text{ext},m}$ , where  $\sum_m M_{\text{ext},m}$  is the total melt fraction extracted during the previous  $m$  extraction episodes. At the beginning (Extended Data Fig. 1), all mantle and plume rocks are assumed to be melt-depleted ( $\sum_m M_{\text{ext},m} = M_0$ ). The rock is considered non-molten (refractory) when the extracted melt fraction is larger than the standard one (that is, when  $\sum_m M_{\text{ext},m} > M_0$ ). If  $M > 0$  for a given marker, then the melt fraction  $M_{\text{ext}} = M$  is extracted and  $\sum_m M_{\text{ext},m}$  is updated. The extracted melt fraction  $M_{\text{ext}}$  is assumed to propagate much faster than the rocks deform<sup>38</sup>. Melt extracted from the mantle and the plume rises and adds to the bottom of the crust as hot melt intrusions (plutons) and to the surface as volcanic rocks<sup>35</sup>. The ratio of volcanic to plutonic rocks was varied as a model parameter in different numerical experiments (Extended Data Table 2). In some experiments, mantle density decrease due to melt extraction<sup>1,41</sup> was taken into account (Extended Data Table 2) according to  $\rho_{0,\text{solid(depl)}} = \rho_{0,\text{solid}}(1 - 0.04 \sum_m M_{\text{ext},m})$ , where  $\rho_{0,\text{solid(depl)}}$  and  $\rho_{0,\text{solid}}$  are the standard densities of depleted and non-depleted solid mantle, respectively (Extended Data Table 1).

Crystallization of magma and melting of the crust are computed from the simple linear batch melting model<sup>31,36</sup>

$$M = \begin{cases} 0 & \text{for } T < T_{\text{solidus}} \\ (T - T_{\text{solidus}}) / (T_{\text{liquidus}} - T_{\text{solidus}}) & \text{for } T_{\text{solidus}} < T < T_{\text{liquidus}} \\ 1 & \text{for } T > T_{\text{liquidus}} \end{cases}$$

where  $M$  is volumetric melt fraction,  $T_{\text{solidus}}$  and  $T_{\text{liquidus}}$  are, respectively, solidus and liquidus temperatures (in kelvin) of the crust at a given pressure  $P$  (in megapascals) (Extended Data Table 1). The effective density of the mafic magma and molten crust is calculated as<sup>36</sup>

$$\rho_{\text{eff}} = \rho_{\text{solid}} \left( 1 - M + M \frac{\rho_{0,\text{molten}}}{\rho_{0,\text{solid}}} \right)$$

where  $\rho_{0,\text{solid}}$  and  $\rho_{0,\text{molten}}$  are the standard densities of solid and molten crust, respectively (Extended Data Table 1), and  $\rho_{\text{solid}}$  is the density of solid crust at given  $P$  and  $T$  computed from  $\rho_{\text{solid}} = \rho_{0,\text{solid}} \times (1 - \alpha(T - 298)) \times (1 + \beta(P - 0.1))$ , where  $\alpha = 3 \times 10^{-5} \text{ K}^{-1}$  and  $\beta = 10^{-5} \text{ MPa}^{-1}$  are thermal expansion and compressibility of the crust, respectively. The effect of latent heating due to equilibrium crystallization of molten rocks is implicit in the heat conservation equation by increasing the effective heat capacity ( $C_{P,\text{eff}}$ ) and the thermal expansion ( $\alpha_{\text{eff}}$ ) of the partially crystallized/molten rocks ( $0 < M < 1$ ), calculated as<sup>36</sup>

$$C_{P,\text{eff}} = C_P + Q_L \left( \frac{\partial M}{\partial T} \right)_{P=\text{const}}$$

$$\alpha_{\text{eff}} = \alpha + \rho \frac{Q_L}{T} \left( \frac{\partial M}{\partial P} \right)_{T=\text{const}}$$

where  $C_p = 1,000 \text{ J kg}^{-1}$  is the heat capacity of the solid crust and  $Q_L = 380 \text{ kJ kg}^{-1}$  is the latent heat of crystallization of the crust<sup>34</sup>.

**Subducted crust dehydration and eclogitization model.** The slab dehydration and mantle hydration model is based on the water markers approach<sup>42</sup>. The equilibrium mineralogical water content is computed for the crust and the mantle as a function of pressure and temperature from thermodynamic data by free-energy minimization<sup>42,43</sup>. In addition, water is present as pore fluid with concentrations of up to 2 wt% in the upper oceanic crust (basalt). The pore water content  $X_{\text{H}_2\text{O}(p)}$  (in weight per cent) decreases linearly from the maximal value of  $X_{\text{H}_2\text{O}(p)} = 2$  wt% at the surface to 0 wt% at a depth of 75 km:  $X_{\text{H}_2\text{O}(p)} = (1 - 0.013\Delta y)X_{\text{H}_2\text{O}(p_0)}$ , where  $\Delta y$  is the depth (0–75 km). The pore water release also mimics effects of low-temperature ( $T < 573 \text{ K}$ ) reactions, which are not included in the thermodynamic database<sup>42</sup>. The oceanic crust dehydrates as it sinks. The timing of  $\text{H}_2\text{O}$  release by dehydration reactions is determined by the physicochemical conditions of the model and the assumption of thermodynamic equilibrium. Water propagates upward into the mantle wedge and induces fluid-fluxed peridotite melting atop subducting slabs. Water propagation is modelled in the form of water markers: dehydration reactions lead to a release of water, the amount of which is stored in a newly generated water marker. Water markers move through the rocks with velocity  $v_{x(\text{water})} = v_x$ ,  $v_{y(\text{water})} = v_y - v_{y(\text{percolation})}$ , where  $v_x$  and  $v_y$  describe the local velocity of the mantle and  $v_{y(\text{percolation})}$  indicates the relative velocity of upward percolation ( $v_{y(\text{percolation})} = 10 \text{ cm yr}^{-1}$  in our experiments)<sup>42</sup>. The marker releases water as soon as it encounters a rock capable of absorbing water by hydration or melting reactions at given pressure and temperature conditions and rock composition<sup>42</sup>. Eclogitization of subducted crust is implemented in a simplified manner, as a linear density increase with pressure from 0% to 16% in the pressure-temperature region between the experimentally determined garnet-in and plagioclase-out phase transitions in basalt<sup>44</sup>.

**Tectonomagmatic plume–lithosphere interaction model.** Melt-induced weakening of the lithosphere above the plume is critical for subduction initiation and implies substantial reduction in the long-term brittle strength of rocks ( $\tau_{\text{yield}}$ ) that are subjected to melt percolation according to the modified Drucker–Prager yield criterion in which the melt pressure effect<sup>14</sup> is considered (equation (1)):  $\tau_{\text{yield}} = C + \phi P \lambda_{\text{melt}}$ . According to this equation, melt pressure plays a critical role, substantially reducing the long-term brittle/plastic strength of the lithosphere through which melt percolates. This effect is especially important for the strong brittle part of the lithosphere, which cannot be weakened by ductile damage<sup>2</sup>. Melt-induced weakening is relevant for high host-rock viscosities ( $\geq 10^{22} \text{ Pa s}$ , values that are found in the cool regions of the lithosphere and crust)<sup>14</sup>. Such weakening corresponds to conditions when shear stress reaches the tensile strength of rocks at nearly equal melt and total pressures<sup>14</sup>. The dominant features of melt transport in this regime are subvertical, planar, sharply localized zones (dykes) in which melt is transported upward from the source region<sup>14</sup>. According to the mechanical energy dissipation balance given in the next section, the long-term lithospheric strength parameter  $\lambda_{\text{melt}}$  is a strain-averaged rather than a time-averaged quantity. Its magnitude is mainly defined by the ratio between melt pressure and lithostatic pressure along dykes during short dyke propagation episodes, which control most of the lithospheric deformation and mechanical energy dissipation. We further quantified the range of expected values for  $\lambda_{\text{melt}}$  by performing 2D numerical hydromechanical experiments on melt-bearing rock deformation as well as seismomechanical experiments on long-term lithospheric deformation assisted by short-term dyke propagation episodes. These numerical experiments showed that the long-term strain-averaged  $\lambda_{\text{melt}}$  is of the order of  $10^{-4}$ – $10^{-2}$  (Extended Data Tables 3, 4, Extended Data Figs 5, 6). The 2D experiments also showed (Extended Data Table 4, Extended Data Fig. 6) that the effective lowering of  $\lambda_{\text{melt}}$  for the deforming lithosphere critically depends on the availability of melt for enabling frequent episodes of dyke propagation through the lithosphere. This implies that arrival of a hotter mantle plume at the base of the lithosphere—characterized by larger melt production rates and dyke injection frequency—should enhance weakening of the overlying lithosphere. Because melt percolation is not directly simulated in our large-scale 3D models, we test different values for the long-term lithospheric weakening factor  $\lambda_{\text{melt}}$  (Extended Data Table 2). We also note that in our tectonomagmatic model we neglected melt-induced thermomechanical erosion of the lithosphere above the plume<sup>45</sup>.

**Physical meaning and magnitude of melt-induced weakening.** The physical meaning of the effective magmatic weakening parameter  $\lambda_{\text{melt}}$  used in the large-scale and long-term subduction initiation models can be interpreted on the basis of simplified mechanical energy dissipation balance. Consider a unit volume of the lithosphere, which deforms slowly during long amagmatic episodes and rapidly during short magmatic episodes. For each computational time step  $\Delta t_{(c)}$  of the large-scale model, time-integrated mechanical energy dissipation ( $H$ ) for this unit volume can be integrated separately for magmatic and amagmatic episodes

$$H = \int \tau_{ij(m)} \dot{\epsilon}_{ij(m)} dt + \int \tau_{ij(a)} \dot{\epsilon}_{ij(a)} dt \quad (2)$$

where  $\tau_{ij(m/a)}$  and  $\dot{\epsilon}_{ij(m/a)}$  are characteristic deviatoric stresses and non-elastic (that is, irreversible viscous-plastic) strain rates, respectively, for magmatic/amagmatic episodes, and the repeated indices  $i, j$  once again denote summation. Evaluating the integrals in equation (2) under the assumption of constant characteristic stresses and strain rates within the respective episodes gives

$$H = 2\tau_{II(m)} \epsilon_{II(m)} + 2\tau_{II(a)} \epsilon_{II(a)} \quad (3)$$

where  $\epsilon_{II(m/a)} = \sqrt{\dot{\epsilon}_{ij(m/a)} \dot{\epsilon}_{ij(m/a)} \Delta t_{(m/a)}^2} / 2$ ,  $\tau_{II(m/a)} = \sqrt{\tau_{ij(m/a)} \tau_{ij(m/a)} / 2}$ ,  $\Delta t_{(c)} = \Delta t_{(m)} + \Delta t_{(a)}$ ,  $\tau_{ij(m/a)} = 2\eta_{\text{eff}(m/a)} \dot{\epsilon}_{ij(m/a)}$  for any unique combination of  $i$  and  $j$ ,  $\tau_{II(m/a)}$  are second invariants of deviatoric stresses,  $\epsilon_{II(m/a)}$  are cumulative strains,  $\Delta t_{(m/a)}$  are cumulative durations and  $\eta_{\text{eff}(m/a)}$  are effective (viscous-plastic) viscosities for the respective episodes. To be mechanically self-consistent, the long-term computational model, which does not resolve rapid magmatic deformation episodes, must predict the correct mechanical energy dissipation as well as the correct total strain for the same computational time step  $\Delta t_{(c)}$ , which thus requires

$$H_{(c)} = \tau_{ij(c)} \dot{\epsilon}_{ij(c)} \Delta t_{(c)} = 2\tau_{II(c)} \epsilon_{II(c)} = H \quad (4)$$

$$\epsilon_{II(c)} = \epsilon_{II(m)} + \epsilon_{II(a)} \quad (5)$$

where  $\epsilon_{II(c)} = \sqrt{\dot{\epsilon}_{ij(c)} \dot{\epsilon}_{ij(c)} \Delta t_{(c)}^2} / 2$ ,  $\tau_{II(c)} = \sqrt{\tau_{ij(c)} \tau_{ij(c)} / 2}$ ,  $H_{(c)}$ ,  $\tau_{ij(c)}$ ,  $\dot{\epsilon}_{ij(c)}$ ,  $\tau_{II(c)}$  and  $\epsilon_{II(c)}$  are respective quantities computed for  $\Delta t_{(c)}$  by the long-term model. By combining equations (3)–(5) we get

$$\tau_{II(c)} = \frac{\epsilon_{II(m)}}{\epsilon_{II(m)} + \epsilon_{II(a)}} \tau_{II(m)} + \frac{\epsilon_{II(a)}}{\epsilon_{II(m)} + \epsilon_{II(a)}} \tau_{II(a)} \quad (6)$$

which defines the modelled second deviatoric stress invariant  $\tau_{II(c)}$  as a strain-averaged product of  $\tau_{II(m)}$  and  $\tau_{II(a)}$ .

We can roughly evaluate possible magnitudes of  $\epsilon_{II(m)}$  and  $\epsilon_{II(a)}$  for a single computational timestep  $\Delta t_{(c)}$ , which typically varies from 30 yr to 300 yr ( $10^9$ – $10^{10}$  s) for our models. In tested models with no melt-induced weakening, lithospheric strain rates  $\dot{\epsilon}_{II(a)} = \sqrt{\dot{\epsilon}_{ij(a)} \dot{\epsilon}_{ij(a)} / 2}$  imposed by the plume vary between  $10^{-17} \text{ s}^{-1}$  and  $10^{-15} \text{ s}^{-1}$  (the effective viscosity of the lithosphere is  $10^{22}$ – $10^{24} \text{ Pa s}$ ). Consequently, maximal accumulated viscous-plastic amagmatic strain  $\epsilon_{II(a)} = \dot{\epsilon}_{II(a)} \Delta t_{(c)}$  for one timestep  $\Delta t_{(c)}$  should be of the order of  $10^{-8}$ – $10^{-5}$ . On the other hand, accumulated elastic amagmatic strain could be released during magmatic episodes as a viscous-plastic magmatic strain  $\epsilon_{II(m)}$ . Elastic strain release should be proportional to the stress level drop between the episodes (that is,  $\Delta \tau = \tau_{II(a)} - \tau_{II(m)}$ ). In tested models with no melt-induced weakening, lithospheric deviatoric stresses imposed by the plume are of the order of  $10^7$ – $10^8 \text{ Pa}$ . This should be notably higher than the brittle/plastic strength of melt-bearing rocks, which is of the order of their tensile strength  $C$  (and varies from 0 to  $5 \times 10^7$  in crystalline rocks)<sup>46</sup>, owing to nearly equal total and melt pressures<sup>14</sup>. By taking a conservative estimate for the elastic stress release  $\Delta \tau$  of the order of  $10^6$ – $10^7 \text{ Pa}$  and using a shear modulus  $G$  of lithospheric rocks (of the order of  $10^{10} \text{ Pa}$ ), we obtain  $\epsilon_{II(m)} = \Delta \tau / G$  of the order of  $10^{-4}$ – $10^{-3}$ , which is at least an order of magnitude larger than the estimated  $\epsilon_{II(a)}$ . We therefore conclude that for the conditions of our long-term numerical experiments, most of the irreversible viscous-plastic strain is accumulated during the magmatic episodes ( $\epsilon_{II(m)} \gg \epsilon_{II(a)}$ ) when the lithospheric rocks are strongly weakened. Therefore, magnitudes of modelled stresses and strains should closely match these for the magmatic episodes (that is,  $\tau_{II(c)} \approx \tau_{II(m)}$ ,  $\epsilon_{II(c)} \approx \epsilon_{II(m)}$ ) irrespective of their short duration (that is, irrespective of  $\Delta t_{(m)} \ll \Delta t_{(a)}$ ).

In our viscous-plastic model,  $\tau_{II(c)}$  in equation (6) is limited by the modified Drucker–Prager yield criterion (equation (1))

$$\tau_{II(c)} \leq C + \phi P \lambda_{\text{melt}} \quad (7)$$

which in combination with equation (6) implies that the long-term model parameters  $\phi$ ,  $C$  and  $\lambda_{\text{melt}}$  are also strain-averaged rather than time-averaged quantities. This, in turn, implies that  $\lambda_{\text{melt}}$  should closely match the magnitude of melt-induced weakening  $\lambda_{\text{melt}(m)}$  during the short magmatic episodes, which control most of the lithospheric deformation and mechanical energy dissipation.

To test these theoretical considerations we quantified strain-averaged  $\lambda_{\text{melt}(m)}$  in deforming melt-bearing rocks<sup>47</sup> (Extended Data Table 3) and calculated the long-term, strain-averaged, melt-induced lithospheric weakening parameter  $\lambda_{\text{melt}}$  (Extended Data Table 4) on the basis of 2D numerical experiments, which are described in the next sections.

**Modelling of melt-bearing rocks deformation.** We performed systematic 2D numerical experiments aimed at directly quantifying strain-averaged  $\lambda_{\text{melt}(m)}$  for

deforming melt-bearing rocks<sup>47</sup> for a broad range of material parameters and deformation conditions.

**Modelling approach.** The numerical hydromechanical 2D Matlab code<sup>48</sup> used for the modelling of melt-bearing rock deformation<sup>47</sup> is based on a combination of a finite-difference method, applied on a staggered Eulerian grid, and a marker-in-cell technique<sup>31,32</sup>. Our coupled hydromechanical system of equations consists of solid and melt mass conservation equations, bulk material momentum conservation equations and Darcy equations for melt migration. These equations are solved on the Eulerian frame under the condition of mutual incompressibility of viscous-plastic solid and low-viscosity melt. Physical properties of the solid matrix (effective viscosity, density, porosity, permeability and so on) are transported by Lagrangian markers that move according to the velocity field interpolated from the fixed grid. Full details of this method, allowing for its reproduction, are provided elsewhere<sup>48</sup>.

**Numerical model design and modelling procedure.** The initial model set-up corresponds to a rectangular Eulerian domain with an aspect ratio of 10 (Extended Data Fig. 5a), which is resolved with a regular rectangular grid of  $301 \times 31$  nodes and contains 144,000 randomly distributed Lagrangian markers. Upper and lower boundaries have constant horizontal velocity, which corresponds to a prescribed simple shear deformation rate. Side boundaries have no slip conditions with the prescribed horizontal velocity changing linearly from the upper to the lower boundary. Lagrangian markers leaving the Eulerian grid are recycled from the other side of the model. These Lagrangian markers correspond to a solid porous melt-bearing rock matrix and are initially assigned by a prescribed value of the volumetric melt content  $\psi$  (porosity) with small random perturbations. At each time step, a numerical solution is obtained on the Eulerian grid for solid matrix velocity  $v_{x(\text{matrix})}$ ,  $v_{y(\text{matrix})}$ , melt pressure  $P_{\text{melt}}$  and total pressure  $P = \psi P_{\text{melt}} + (1 - \psi)P_{\text{matrix}}$  (refs 14, 48). The local melt content evolves during the numerical experiments in accordance to the implemented viscous-plastic compaction equation<sup>48</sup>. The experiments are run for several hundred time steps until melt bands<sup>47</sup> are formed, which localize deformation and energy dissipation and have strongly reduced  $\lambda_{\text{melt}(m)} = 1 - P_{\text{melt}}/P$  signatures (Extended Data Fig. 5a–g). After near-steady-state values of strain-averaged  $\lambda_{\text{melt}(m)}$  are reached (Extended Data Fig. 5h), they are averaged for 200 time steps (Extended Data Table 3).

**Viscous-plastic rheology and permeability of melt-bearing rock matrix.** The viscous and brittle (plastic) properties of the matrix are implemented via evaluation of the effective viscosity. Ductile viscosity of the matrix depends exponentially<sup>47</sup> on the melt content  $\psi$ :  $\eta_{\text{ductile}} = \eta_0 e^{-a\psi}$ , where  $\eta_0$  is a reference viscosity and  $a = 28$  is an experimentally derived melt-weakening coefficient<sup>47,49,50</sup>. The ductile rheology is combined with a brittle (plastic) rheology to yield an effective viscous-plastic rheology using the following upper limit for the ductile viscosity, which depends on the local melt content as well as on the difference between total pressure  $P$  and melt pressure  $P_{\text{melt}}$

$$\eta_{\text{ductile}} \leq \frac{C(1 - \psi) + \phi(P - P_{\text{melt}})}{2\dot{\epsilon}_{\text{II}(s)}} \quad (8)$$

where  $\dot{\epsilon}_{\text{II}(s)} = \sqrt{\dot{\epsilon}_{ij}(s)\dot{\epsilon}_{ij}(s)}/2$  is the second strain rate invariant for the solid matrix,  $\phi < 1$  for confined fractures<sup>46</sup> when  $P > P_{\text{melt}}$  (Extended Data Table 3) and  $\phi = 1$  for tensile fractures<sup>36</sup> when  $P < P_{\text{melt}}$ . Solid matrix permeability  $k$  depends on the volumetric melt content as<sup>48</sup>  $k = k_0(\psi/\psi_0)^3$ , where  $k_0$  and  $\psi_0 = 0.01$  are the reference permeability and melt content, respectively. The melt content  $\psi$  changes with time  $t$  in accordance with the viscous-plastic compaction equation<sup>48</sup>

$$\frac{1}{\psi} \frac{\partial \psi}{\partial t} = - \frac{P - P_{\text{melt}}}{\eta_{\text{bulk}}}$$

where  $\eta_{\text{bulk}} = \eta_{\text{ductile}}/\psi$  is the bulk viscosity of the matrix and  $\eta_{\text{ductile}}$  is corrected in accordance with equation (8).

**Results of experiments.** The hydromechanical numerical experiments show that the strain-averaged  $\lambda_{\text{melt}(m)}$  is of the order of  $10^{-4}$ – $10^{-2}$  (Extended Data Table 3). They also show that  $\lambda_{\text{melt}(m)}$  in deforming melt-bearing rocks correlates positively with the tensile strength  $C$  and friction coefficient  $\phi$ , and negatively with confining pressure  $P_{\text{conf}}$ . The magnitude of difference between  $P$  and  $P_{\text{melt}}$  is, however, nearly independent of  $P_{\text{conf}}$  (Extended Data Table 3). Other parameters have less influence on  $\lambda_{\text{melt}(m)}$ : positive correlation is documented for matrix viscosity and deformation rate, whereas negative correlation is found for matrix permeability, melt viscosity and volumetric melt fraction (Extended Data Table 3).

**Modelling of melt-induced weakening of the lithosphere.** We performed systematic 2D seismomechanical numerical experiments aimed at directly quantifying long-term (tens to hundreds of years) strain-averaged  $\lambda_{\text{melt}}$  in deforming visco-elasto-plastic lithosphere subjected to short-term (hours) dyke propagation episodes. Dykes are a dominant mode of low-viscosity mafic-melt transport through the high-viscosity lithosphere<sup>14,51</sup> and their propagation and emplacement is

associated with rapid rock deformation and related seismicity<sup>52,53</sup>. Dyke swarms are a characteristic feature of the lithosphere above mantle plumes<sup>22,23</sup>.

**Modelling approach.** The seismomechanical version of the I2ELVIS code<sup>54</sup> used for the modelling of lithospheric deformation assisted by frequent episodes of dyke propagation<sup>52,53</sup> is based on a combination of a finite-difference method, applied on a staggered Eulerian grid, and a marker-in-cell technique<sup>31,32</sup>. This code allows for modelling both slow interseismic deformation and rapid localized deformation associated with seismogenesis<sup>54</sup>. A coupled seismomechanical system of equations consists of mass and momentum conservation equations, which account for inertial forces and incompressible visco-elasto-plastic rheology of rocks. Full details of this method, allowing for its reproduction, are provided elsewhere<sup>54</sup>.

**Numerical model design and modelling procedure.** The initial model set-up corresponds to a rectangular  $3 \text{ km} \times 3 \text{ km}$  lithospheric domain (Extended Data Fig. 6a), which is resolved with a regular rectangular grid of  $101 \times 101$  nodes and contains 160,000 randomly distributed Lagrangian markers. All boundaries are free slip and have constant normal velocity corresponding to a prescribed pure shear strain rate (extension direction is horizontal). Material properties inside the domain correspond to rheologically strong lithosphere with high shear modulus ( $G_{\text{lith}} = 67 \text{ GPa}$ )<sup>34</sup>, viscosity ( $\eta_{\text{lith}}$ ) and brittle strength  $\tau_{\text{yield}(\text{lith})} = C + \phi P$  (Extended Data Table 4). The lithosphere deforms predominantly elastically during relatively long (up to several years) amagmatic episodes, which are alternated with very short (a few hours) magmatic dyke propagation episodes.

During each magmatic episode, a thin rheologically weak planar dyke is prescribed at a new position with the smallest lithospheric strength excess relative to the local deviatoric stress. The dyke has its orientation randomly chosen in the range of  $75^\circ$ – $105^\circ$  to the extension direction<sup>55</sup> and is characterized by a lowered brittle/plastic strength  $\tau_{\text{yield}(\text{dyke})} = C + \phi P \lambda_{\text{melt}(m)}$  corresponding to the assumed characteristic  $\lambda_{\text{melt}(m)}$  for melt-bearing dyke rocks (Extended Data Tables 3, 4). As a result, rapid visco-elasto-plastic deformation localizes along the rheologically weak dyke, thus releasing part of the elastic stresses accumulated by surrounding elastic lithospheric rocks during preceding amagmatic episodes. At the end of the magmatic episode, the low strength of the dyke is increased to high ambient strength of the lithosphere.

The duration of amagmatic episodes  $t_a$  is varied in different experiments and is defined according to the characteristic dyke intrusion frequency ( $I_f$ )<sup>56</sup>

$$I_f = \frac{1}{t_0} = \frac{2G_{\text{lith}} \dot{\epsilon}_{\text{extension}}}{C}$$

where  $\dot{\epsilon}_{\text{extension}}$  is the imposed lithospheric extension strain rate and  $t_0$  is a characteristic time of extension needed to overcome ambient tensile strength of the lithosphere  $C$ . We tested a range of  $t_a$  values (Extended Data Table 4) to take into account possible effects of incomplete lithospheric stress release during dyke propagation events ( $t_a < t_0$ ) and unavailability of melt needed for dyke propagation ( $t_a > t_0$ ).

The model is run for 10,000 time steps, which corresponds to 100 dyke emplacement episodes. In accordance with equations (2) and (4), effective long-term visco-plastic strength of the lithosphere ( $\tau_{\text{eff}}$ ) is computed from the integrated viscous-plastic strain and energy dissipation as

$$\tau_{\text{eff}} = \frac{\iint \tau_{ij} \dot{\epsilon}_{ij}(\text{vp}) dV dt}{2 \iint \sqrt{\dot{\epsilon}_{ij}(\text{vp}) \dot{\epsilon}_{ij}(\text{vp})/2} dV dt}$$

where  $\dot{\epsilon}_{\text{II}(\text{vp})}$  is viscous-plastic strain rate,  $\tau_{ij}$  is the deviatoric stress and  $\iint dV dt$  corresponds to integration over the model volume and time. Effective long-term  $\lambda_{\text{melt}}$  for the lithosphere is then computed in accordance with equation (7) as

$$\lambda_{\text{melt}} = \frac{\tau_{\text{eff}} - C}{P_{\text{lith}} \phi}$$

where  $P_{\text{lith}}$  is the average pressure in the lithosphere prescribed in the experiment (Extended Data Table 4). For comparison with  $\tau_{\text{eff}}$ , the long-term average volumetric lithospheric stress in the lithosphere ( $\tau_{\text{vol}}$ ) is calculated as

$$\tau_{\text{vol}} = \frac{1}{tV} \iint \tau_{\text{II}} dV dt$$

where  $\tau_{\text{II}}$  is the second deviatoric stress invariant,  $V$  is model volume and  $t$  is time at the end of the experiment.

**Results of experiments.** The numerical experiments show (Extended Data Table 4, Extended Data Fig. 6) that both viscous-plastic strain  $\epsilon_{\text{II}(\text{vp})}$  and energy dissipation are localized along dykes (Extended Data Fig. 6a, c, d) and restricted to the short dyke propagation episodes (Extended Data Fig. 6e, f). As a result, effective long-term visco-plastic strength of the deforming lithosphere subjected to frequent dyke

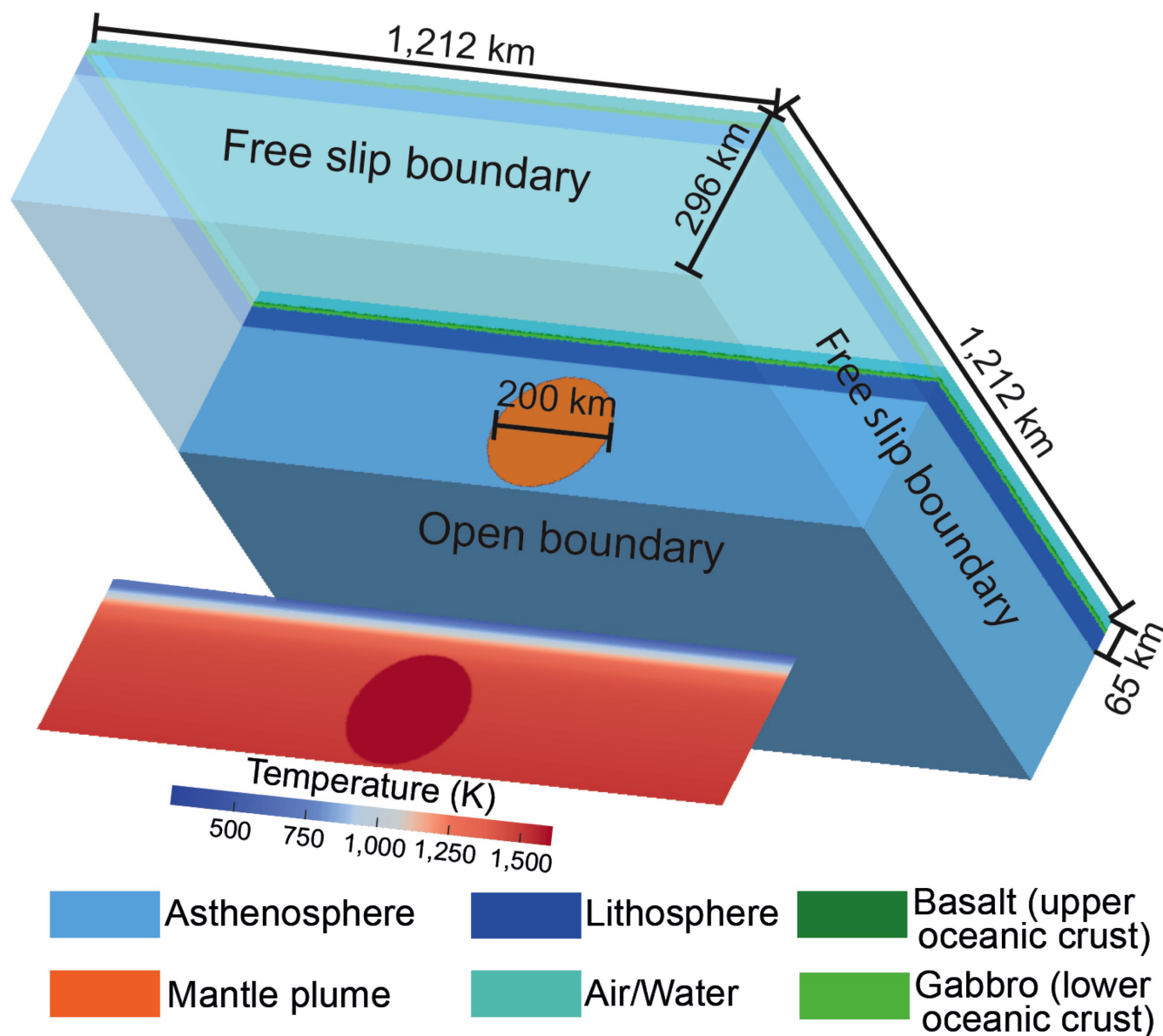


propagation  $\tau_{\text{eff}}$  is much lower than its ambient viscous-plastic strength  $\tau_{\text{yield(lith)}}$ . It is also notably lower than the long-term average volumetric lithospheric stress  $\tau_{\text{vol}}$  (Extended Data Fig. 6g), because most of the strain is accumulated along rheologically weak, propagating, melt-bearing dykes characterized by reduced deviatoric stresses compared to surrounding solid lithospheric rocks (Extended Data Fig. 6b). Experiments also show that the effective long-term  $\lambda_{\text{melt}}$  for the lithosphere deforming under conditions of frequent dyke propagation (Extended Data Fig. 6h) is of the order of  $10^{-3}$ – $10^{-2}$  (Extended Data Table 4), which is comparable to the strain-averaged  $\lambda_{\text{melt(m)}}$  of melt-bearing rocks (Extended Data Table 3), and thus the experiments support our simple theoretical analyses based on energy dissipation. The experiments also show that  $\lambda_{\text{melt}}$  positively correlates with the duration of amagmatic episodes  $t_a$  and that frequent  $\left(t_a \leq \frac{10^6}{2G_{\text{lith}}\epsilon_{\text{extension}}}\right)$  propagation

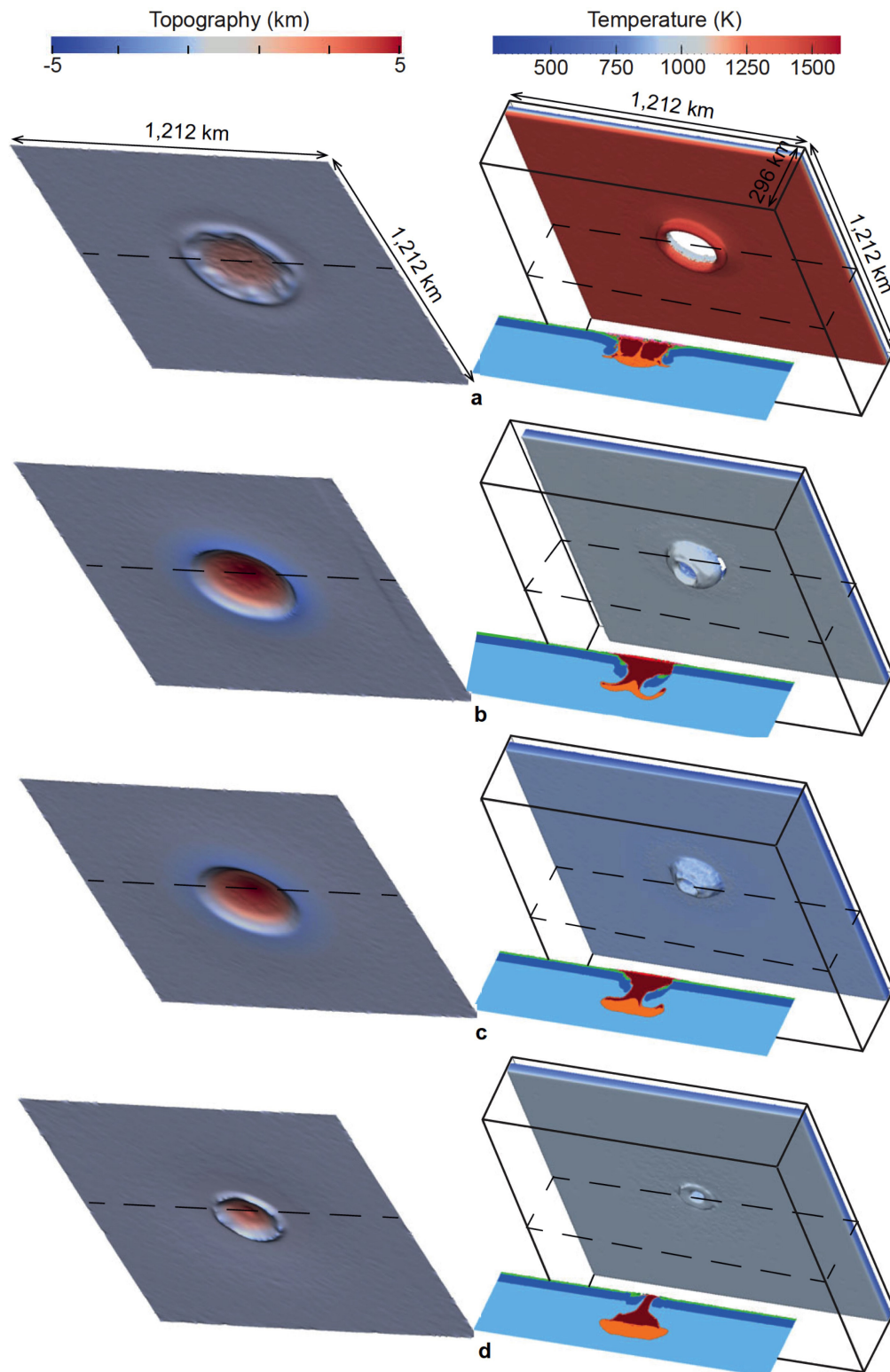
of dykes (and thus larger melt flux from the plume) is needed to substantially lower the effective strength of the deforming lithosphere. Some other parameters also have influence on  $\lambda_{\text{melt}}$ : positive correlation is documented for strain rate and  $\lambda_{\text{melt(m)}}$  of dykes, whereas negative correlation is found for dyke length, lithospheric viscosity, average pressure in the lithosphere  $P_{\text{lith}}$ , friction coefficient  $\phi$  and tensile strength  $C$  (Extended Data Table 4).

**Code availability.** Open-source software ParaView (<http://www.paraview.org>) was used for the simulations shown in Figs 1, 3, 4 and Extended Data Figs 1–4. We have opted not to make the numerical modelling codes associated with this paper available.

31. Gerya, T. V. *Introduction to Numerical Geodynamic Modelling* (Cambridge Univ. Press, 2010).
32. Gerya, T. V. & Yuen, D. A. Robust characteristics method for modelling multiphase visco-elasto-plastic thermo-mechanical problems. *Phys. Earth Planet. Inter.* **163**, 83–105 (2007).
33. Schmeling, H. *et al.* A benchmark comparison of spontaneous subduction models: towards a free surface. *Phys. Earth Planet. Inter.* **171**, 198–223 (2008).
34. Turcotte, D. L. & Schubert, G. *Geodynamics* (Cambridge Univ. Press, 2002).
35. Vogt, K., Gerya, T. V. & Castro, A. Crustal growth at active continental margins: numerical modeling. *Phys. Earth Planet. Inter.* **192–193**, 1–20 (2012).
36. Gerya, T. V. Three-dimensional thermomechanical modeling of oceanic spreading initiation and evolution. *Phys. Earth Planet. Inter.* **214**, 35–52 (2013).
37. Katz, R. F. Porosity-driven convection and asymmetry beneath mid-ocean ridges. *Geochem. Geophys. Geosyst.* **11**, Q0AC07 (2010).
38. Connolly, J. A. D., Schmidt, M. W., Solferino, G. & Bagdassarov, N. Permeability of asthenospheric mantle and melt extraction rates at mid-ocean ridges. *Nature* **462**, 209–212 (2009).
39. Katz, R. F., Spiegelman, M. & Langmuir, C. H. A new parameterization of hydrous mantle melting. *Geochem. Geophys. Geosyst.* **4**, 1073 (2003).
40. Nikolaeva, K., Gerya, T. V. & Connolly, J. A. D. Numerical modelling of crustal growth in intraoceanic volcanic arcs. *Phys. Earth Planet. Inter.* **171**, 336–356 (2008).
41. Schutt, D. L. & Leshner, C. E. Effects of melt depletion on the density and seismic velocity of garnet and spinel lherzolite. *J. Geophys. Res.* **111**, B05401 (2006).
42. Gerya, T. V. & Meilick, F. I. Geodynamic regimes of subduction under an active margin: effects of rheological weakening by fluids and melts. *J. Metamorph. Geol.* **29**, 7–31 (2011).
43. Connolly, J. A. D. Computation of phase equilibria by linear programming: a tool for geodynamic modeling and its application to subduction zone decarbonation. *Earth Planet. Sci. Lett.* **236**, 524–541 (2005).
44. Ito, K. & Kennedy, G. C. in *The Structure and Physical Properties of the Earth's Crust* (ed. Heacock, J. G.) 303–314 (Geophysical Monograph Series, Vol. 14, AGU, 1971).
45. Sobolev, S. V. *et al.* Linking mantle plumes, large igneous provinces and environmental catastrophes. *Nature* **477**, 312–316 (2011).
46. Byerlee, J. Friction of rocks. *Pure Appl. Geophys.* **116**, 615–626 (1978).
47. Katz, R. F., Spiegelman, M. & Holtzman, B. The dynamics of melt and shear localization in partially molten aggregates. *Nature* **442**, 676–679 (2006).
48. Dymkova, D. & Gerya, T. Porous fluid flow enables oceanic subduction initiation. *Geophys. Res. Lett.* **40**, 5671–5676 (2013).
49. Mei, S., Bai, W., Hiraga, T. & Kohlstedt, D. Influence of melt on the creep behavior of olivine-basalt aggregates under hydrous conditions. *Earth Planet. Sci. Lett.* **201**, 491–507 (2002).
50. Kelemen, P., Shimizu, N. & Salters, V. Extraction of mid-ocean-ridge basalt from the upwelling mantle by focused flow of melt in dunite channels. *Nature* **375**, 747–753 (1995).
51. Rubin, A. M. Dykes vs. diapirs in viscoelastic rock. *Earth Planet. Sci. Lett.* **119**, 641–659 (1993).
52. Pedersen, R., Sigmundsson, F. & Einarsson, P. Controlling factors on earthquake swarms associated with magmatic intrusions; constraints from Iceland. *J. Volcanol. Geotherm. Res.* **162**, 73–80 (2007).
53. Sigmundsson, F. *et al.* Intrusion triggering of the 2010 Eyjafjallajökull explosive eruption. *Nature* **468**, 426–430 (2010).
54. van Dinther, Y. *et al.* The seismic cycle at subduction thrusts: 2. Dynamic implications of geodynamic simulations validated with laboratory models. *J. Geophys. Res.* **118**, 1502–1525 (2013).
55. Gudmundsson, A. Emplacement and arrest of sheets and dykes in central volcanoes. *J. Volcanol. Geotherm. Res.* **116**, 279–298 (2002).
56. Gudmundsson, A. Emplacement of dykes, sills and crustal magma chambers at divergent plate boundaries. *Tectonophysics* **176**, 257–275 (1990).
57. Clauser, C. & Huenges, E. in *Rock Physics & Phase Relations: A Handbook of Physical Constants* (ed. Ahrens, T. J.) 105–126 (Am. Geophys. Union, 1995).
58. Hess, P. C. *Origin of Igneous Rocks* (Harvard Univ. Press, 1989).
59. Schmidt, M. W. & Poli, S. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet. Sci. Lett.* **163**, 361–379 (1998).
60. Ranalli, G. *Rheology of the Earth* 2nd edn (Chapman & Hall, 1995).



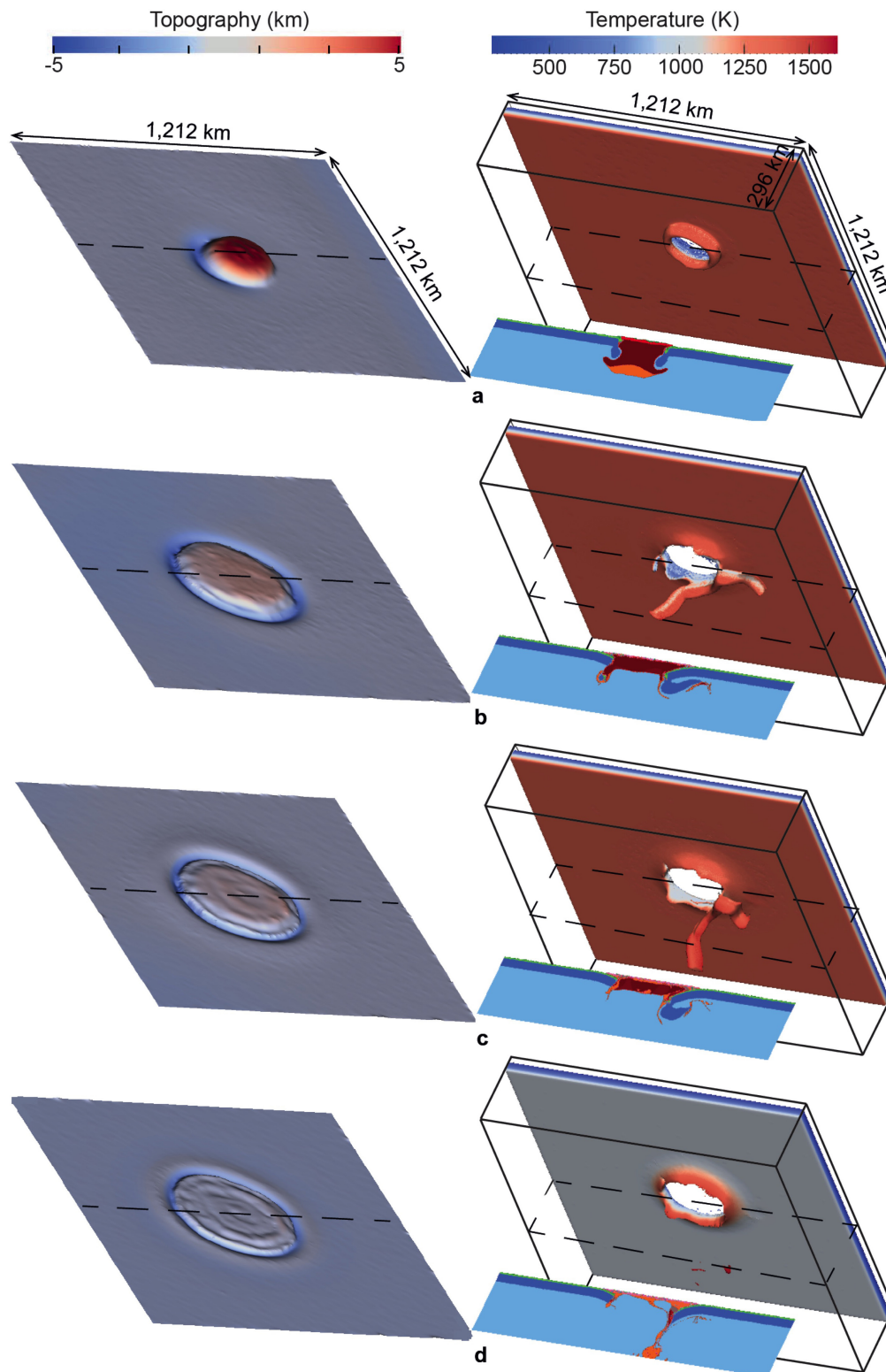
**Extended Data Figure 1 | Numerical model design and boundary conditions.** See Methods for details. 2D cross-section through the centre of the model shows the initial temperature distribution. Colour code for different materials is shown at the bottom of the figure.



**Extended Data Figure 2 | Influence of the oceanic plate age for plume-induced subduction initiation.** Compare with the reference model shown in Fig. 1. **a**, 10-Myr-old plate (model, 'M1' at 2.55 Myr). **b**, 60-Myr-old plate (model, 'M2' at 0.47 Myr). **c**, 80-Myr-old plate (model, 'M3' at

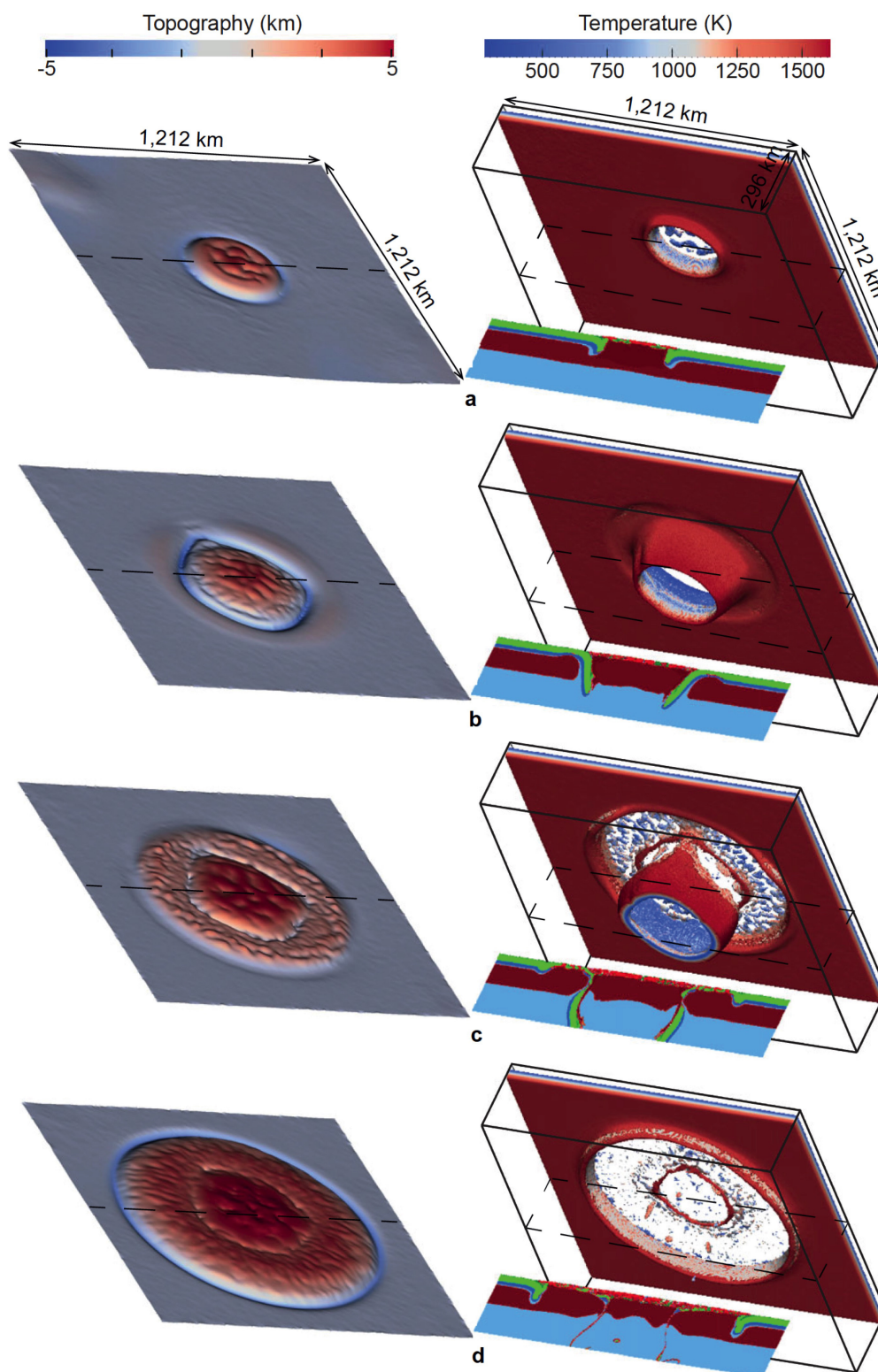
0.42 Myr). **d**, 120-Myr-old plate (model, 'M4' at 0.42 Myr). Dashed lines indicate positions of the 2D cross-sections shown in the right column. Colour code is the same as that in Fig. 1. See Extended Data Table 2 for details of the models.





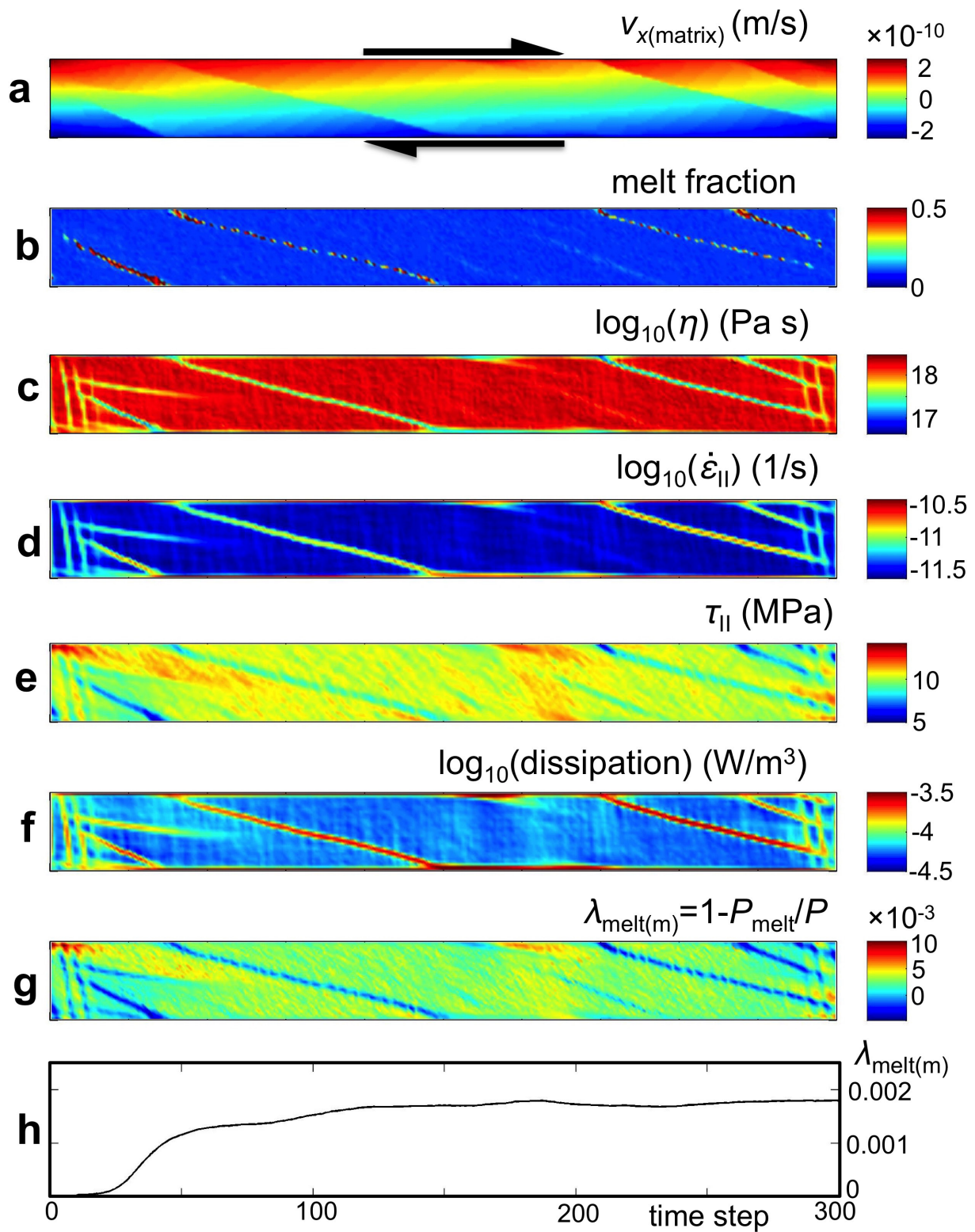
**Extended Data Figure 3 | Freezing of plume-induced subduction due to increased strength of upper (basaltic) oceanic crust.** Compare with the reference model shown in Fig. 1. The model used for the simulations was 'bsay'; see Extended Data Table 2. **a**, Oceanic plateau development

(at 0.11 Myr). **b**, Subduction initiation and tearing of the circular slab (at 1.15 Myr). **c**, Freezing of subduction and detachment of lithospheric drips from the slab edges (at 4.67 Myr). **d**, Cooling of the plume and thermal relaxation of the lithosphere (at 29.65 Myr).



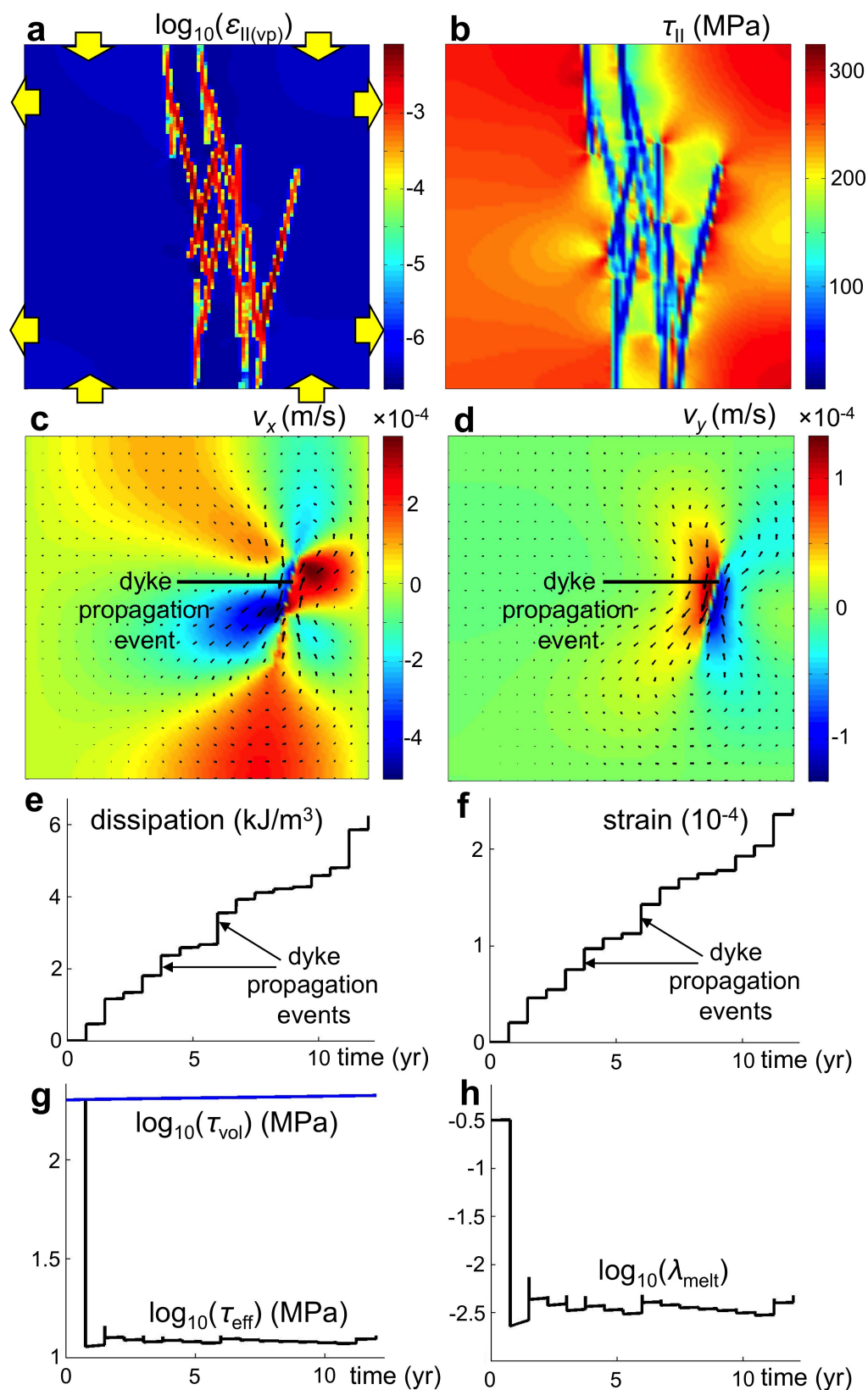
**Extended Data Figure 4 | Dynamics of plume-induced lithospheric drips for 20-Myr-old plate with 30-km-thick crust formed under hotter mantle temperature conditions.** The model used for the simulations was 'bsar'; see Extended Data Table 2. **a**, Oceanic plateau development

(at 0.05 Myr). **b**, Formation of a circular eclogitic crustal drip at the plateau margins (at 0.26 Myr). **c**, Detachment of the circular eclogitic drip (at 0.34 Myr). **d**, Broadening of the plateau and nucleation of the subsequent circular eclogitic drip (at 0.46 Myr).



**Extended Data Figure 5 | 2D hydromechanical numerical model of melt-bearing visco-plastic rock deformation.** See Methods for details. **a–h**, Results for time step 300 are shown. Model domain size is 500 m  $\times$  50 m. Parameters of the numerical experiment correspond to the model ‘L27’ in Extended Data Table 3.





**Extended Data Figure 6 | 2D seismomechanical numerical model of visco-elasto-plastic lithosphere deformation assisted by frequent episodes of dyke propagation.** See Methods for details. **a–h**, Results

for time step 1,555 (propagation of dyke number 16) are shown. Model domain size is 3,000 m  $\times$  3,000 m. Parameters of the numerical experiment correspond to the model 'D65' in Extended Data Table 4.

Extended Data Table 1 | Physical properties of rocks used in numerical experiments

Material	$\rho_0$ , (kg/m <sup>3</sup> )	Thermal conductivity <sup>57</sup> , (W/m/K, at $T_K$ )	$T_{\text{solidus}}^{58,59}$ (K, at $P_{\text{MPa}}$ )	$T_{\text{liquidus}}^{58,59}$ (K, at $P_{\text{MPa}}$ )	Flow law <sup>60</sup>
Upper oceanic crust (basalt)	3000 (solid) 2900 (molten)	$1.18+474/(T+77)$	$973 - 70400/(P + 354) + 77800000/(P + 354)^2$ at $P < 1600$ MPa, $935 + 0.0035P + 0.0000062P^2$ at $P > 1600$ MPa	$1423 + 0.105P$	wet quartzite, $A_D = 1.97 \times 10^{17}$ Pa <sup>n</sup> s, $n=2.3$ , $E=154000$ J/mol, $V=0$ J/mol/MPa, $\sigma_{\text{cr}}=3 \times 10^4$ Pa, $C=1$ MPa, $\phi_0=0$
Lower oceanic crust (gabbro)	3000 (solid) 2900 (molten)	$1.18+474/(T+77)$	$1327+0.091P$	$1423 + 0.105P$	plagioclase An <sub>75</sub> , $A_D = 4.80 \times 10^{22}$ Pa <sup>n</sup> s, $n=3.2$ , $E=238000$ J/mol, $V=0$ J/mol/MPa, $\sigma_{\text{cr}}=3 \times 10^4$ Pa, $C=1$ MPa, $\phi_0=0.1$
Newly formed crust	3000 (solid) 2900 (molten)	$1.18+474/(T+77)$	$973 - 70400/(P + 354) + 77800000/(P + 354)^2$ at $P < 1600$ MPa, $935 + 0.0035P + 0.0000062P^2$ at $P > 1600$ MPa	$1423 + 0.105P$	plagioclase An <sub>75</sub> , $A_D = 4.80 \times 10^{22}$ Pa <sup>n</sup> s, $n=3.2$ , $E=238000$ J/mol, $V=0$ J/mol/MPa, $\sigma_{\text{cr}}=3 \times 10^4$ Pa, $C=1$ MPa, $\phi_0=0.2$
Lithosphere - asthenosphere dry mantle	3300 (solid) 2900 (molten)	$0.73+1293/(T+77)$	*	*	dry olivine, $A_D = 3.98 \times 10^{16}$ Pa <sup>n</sup> s, $n=3.5$ , $E=532000$ J/mol, $V=8$ J/mol/MPa, $\sigma_{\text{cr}}=3 \times 10^4$ Pa, $C=1$ MPa, $\phi_0=0.2$
Lithosphere - asthenosphere wet mantle	3300 (solid) 2900 (molten)	$0.73+1293/(T+77)$	*	*	wet olivine, $A_D = 5.01 \times 10^{20}$ Pa <sup>n</sup> s, $n=4.0$ , $E=470000$ J/mol, $V=8$ J/mol/MPa, $\sigma_{\text{cr}}=3 \times 10^7$ Pa, $C=1$ MPa, $\phi_0=0.2$

Physical properties are taken from refs 57–60 as indicated.

\*Melting model from ref. 39 is used (Methods).

Extended Data Table 2 | Conditions and results of 3D numerical experiments for plume-induced subduction initiation

Model	Mantle potent. temp. (K)	Plume diam. (km)	Plume temp. (K)	Plume rheol.	Plate thickn. °: Cooling age (km:Ma)	Crust thickn. (km)	$\phi_0$ of mantle rocks	$\lambda_{\text{melt}}$	New crust volcan. : pluton. (%)	Results
bsab	1573	200	1858	wet	50:20	8	0.2	0.01	100 : 0	self-sustained subduction
bsan <sup>a</sup>	1573	200	1911	wet	50:20	8	0.2	0.03	20 : 80	plate thinning, oceanic plateau formation
bsar	1773	200	2058	wet	40:20	30	0.2	0.01	20 : 80	circular lithospheric drips
bsas	1573	200	1858	wet	50:20	8	0.2	0.01	20 : 80	self-sustained subduction
bsasb	1573	200	1858	wet	50:20	8	0.2	0.03	20 : 80	oceanic plateau formation
bsay <sup>f</sup>	1573	200	1858	wet	50:20	8	0.2	0.01	20 : 80	subduction initiation and freezing
bsayea	1573	200	1858	wet	50:20	8	0.2	0.01	20 : 80	self-sustained subduction, spreading
bsayeb <sup>c</sup>	1573	200	1858	wet	50:20	8	0.2	0.01	20 : 80	subduction initiation and freezing
bsayee <sup>bc</sup>	1573	200	1858	dry	50:20	8	0.2	0.01	20 : 80	subduction initiation and freezing
bsayef <sup>bce</sup>	1573	200	1858	dry	50:20	8	0.2	0.01	20 : 80	subduction initiation and freezing
bsayeg <sup>bc</sup>	1573	200	1858	wet	60:30	8	0.2	0.01	20 : 80	subduction initiation
bsayeh <sup>bc</sup>	1573	200	1858	dry	60:30	8	0.2	0.01	20 : 80	subduction initiation and freezing
bsayei <sup>bcd</sup>	1573	200	1858	dry	60:30	8	0.2	0.01	20 : 80	subduction initiation and freezing
bsayej <sup>bcd</sup>	1573	200	1858	dry	65:40	8	0.2	0.01	20 : 80	self-sustained subduction
bsayek <sup>bcd</sup>	1573	200	1858	dry	95:80	8	0.2	0.01	20 : 80	subduction initiation
bsayg	1773	200	2058	wet	75:80	8	0.2	0.01	20 : 80	self-sustained subduction
bsayh	1773	200	2058	wet	75:80	20	0.2	0.01	20 : 80	self-sustained subduction
bsayha	1773	200	2058	wet	75:80	20	0.6	0.01	20 : 80	subduction initiation
bsayhb <sup>b</sup>	1773	200	2058	dry	75:80	20	0.2	0.01	20 : 80	subduction initiation
bsayhd <sup>c</sup>	1773	200	2058	wet	75:80	20	0.2	0.01	20 : 80	subduction initiation
bsayhg <sup>bc</sup>	1773	200	2058	dry	75:80	20	0.2	0.01	20 : 80	subduction initiation
bsayhh <sup>bcd</sup>	1773	200	2058	dry	75:80	20	0.2	0.01	20 : 80	self-sustained subduction
bsayi	1773	200	2058	wet	75:80	30	0.2	0.01	20 : 80	self-sustained subduction
bsayn	1573	200	1858	wet	80:60	8	0.2	0.01	20 : 80	plate thinning, oceanic plateau formation and collapse
bsayna	1573	200	1858	dry	80:60	8	0.2	0.01	20 : 80	self-sustained subduction
bsaynb	1573	200	1858	dry	95:80	8	0.2	0.01	20 : 80	self-sustained subduction
bsaync	1573	200	1858	wet	65:40	8	0.2	0.01	20 : 80	subduction initiation and freezing
bsaynd	1573	200	1858	dry	65:40	8	0.2	0.01	20 : 80	plate thinning, oceanic plateau formation and collapse
bsaynh	1573	200	1858	wet	65:40	8	0.4	0.01	20 : 80	oceanic plateau formation
bsayni	1573	200	1858	wet	65:40	8	0.6	0.01	20 : 80	oceanic plateau formation
bsaynj	1573	200	1858	wet	65:40	8	0.3	0.01	20 : 80	subduction initiation
bsaynk	1573	200	1858	wet	65:40	8	0.4	0.003	20 : 80	subduction initiation and freezing
bsaynl	1573	200	1858	wet	65:40	8	0.6	0.003	20 : 80	plate thinning, oceanic plateau formation and collapse
bsaynm	1573	200	1858	wet	65:40	8	0.3	0.003	20 : 80	plate thinning, oceanic plateau formation and collapse
bsaypa	1773	200	2058	dry	75:80	20	0.2	0.01	20 : 80	self-sustained subduction
bsayt	1573	200	1858	wet	50:20	8	0.6	0.01	20 : 80	plate thinning, oceanic plateau formation and collapse
bsayx	1573	200	1858	wet	50:20	8	0.6	0.003	20 : 80	subduction initiation and freezing
bsayy	1573	200	1858	dry	50:20	8	0.2	0.01	20 : 80	subduction initiation and freezing
bsayya <sup>b</sup>	1573	200	1858	dry	50:20	8	0.2	0.01	20 : 80	self-sustained subduction, spreading
bsba	1573	200	1858	wet	50:20	8	0.2	0.03	20 : 80	plate thinning, oceanic plateau formation
M1	1573	200	1858	wet	35:10	8	0.2	0.01	20 : 80	plate thinning, oceanic plateau formation and collapse
M2	1573	200	1858	wet	80:60	8	0.2	0.01	20 : 80	subduction initiation
M3	1573	200	1858	wet	95:80	8	0.2	0.01	20 : 80	subduction initiation
M4	1573	200	1858	wet	110:120	8	0.2	0.01	20 : 80	plate thinning, oceanic plateau formation

<sup>a</sup>Long-lived plume conduit is prescribed.<sup>b</sup>High velocity of fluid percolation from the slab ( $V_{\text{f(percolation)}} = 100 \text{ cm yr}^{-1}$ ).<sup>c</sup>Compositional buoyancy<sup>1,41</sup> of the depleted mantle and plume is taken into account (Methods) and pre-existing self-consistent mantle depletion profile is defined, which corresponds to dry melting<sup>39</sup> along the respective mantle potential temperature profile with an adiabatic thermal gradient of  $0.5 \text{ K km}^{-1}$ .<sup>d</sup>Wet olivine rheology of the asthenosphere ( $A_0 = 5.01 \times 10^{20} \text{ Pa}^n \text{ s}$ ,  $n = 4.0$ ,  $E = 470,000 \text{ J mol}^{-1}$ ,  $V = 8 \text{ J mol}^{-1} \text{ MPa}^{-1}$ ,  $\sigma_{\text{cr}} = 3 \times 10^4 \text{ Pa}$ ,  $C = 1 \text{ MPa}$ ,  $\phi_0 = 0.2$ ).<sup>e</sup>Plate thickness corresponds to the depth of the 1,373-K isotherm.<sup>f</sup>Increased strength of upper (basaltic) oceanic crust ( $\phi_0 = 0.1$ ).



Extended Data Table 3 | Conditions and results of 2D hydromechanical numerical experiments for melt-bearing rock deformation

Run	$\psi$	$\eta_0$ (Pa s)	$\eta_{\text{melt}}$ (Pa s)	$P_{\text{conf}}$ (Pa)	$k_0$ (m <sup>2</sup> )	Strain rate (1/s)	$C$ (Pa)	$\phi$	$P-P_{\text{melt}}$ (Pa)	$\lambda_{\text{melt(m)}}$
L2	0.05	3e+12	10	1e+09	1e-17	1e-04	4e+07	0.6	8.3e+06	0.0083
L3	0.05	3e+12	10000	1e+09	1e-17	1e-04	4e+07	0.6	3.2e+06	0.0032
L5	0.05	3e+14	10	1e+09	1e-17	1e-04	4e+07	0.6	1.3e+06	0.0013
L6	0.05	1e+22	10	1e+09	1e-17	1e-12	4e+07	0.6	2.7e+07	0.0270
L7	0.05	1e+22	10	1e+09	1e-13	1e-13	1e+07	0.6	6.7e+06	0.0067
L8	0.05	1e+22	10	1e+09	1e-13	1e-13	1e+06	0.6	8.0e+05	0.0008
L9	0.05	1e+22	10	3e+09	1e-13	1e-13	1e+07	0.6	7.5e+06	0.0025
L10	0.05	1e+22	10	1e+09	1e-13	1e-13	1e+07	0.2	1.2e+06	0.0012
L13	0.01	1e+22	10	1e+09	1e-13	1e-13	1e+07	0.6	1.8e+07	0.0178
L14	0.05	1e+22	10	1e+09	1e-17	1e-13	1e+07	0.6	6.2e+06	0.0062
L15	0.1	1e+22	10	1e+09	1e-13	1e-13	1e+07	0.6	3.1e+07	0.0310
L18	0.1	3e+21	10	1e+09	1e-13	1e-13	1e+07	0.6	1.7e+06	0.0017
L19	0.1	1e+23	10	1e+09	1e-13	1e-13	1e+07	0.6	4.4e+06	0.0044
L20	0.1	1e+24	10	1e+09	1e-13	1e-13	1e+07	0.6	2.9e+06	0.0029
L21	0.1	1e+22	10	1e+09	1e-13	1e-12	1e+07	0.6	4.4e+06	0.0044
L22	0.1	1e+24	10	1e+09	1e-11	1e-11	1e+06	0.2	1.6e+05	0.0002
L23	0.1	1e+24	10	1e+09	1e-11	1e-11	3e+07	0.6	7.3e+06	0.0073
L24	0.1	1e+24	10	1e+09	1e-11	1e-11	1e+07	0.6	2.4e+06	0.0024
L25	0.1	1e+20	10	1e+09	1e-11	1e-11	1e+07	0.6	3.4e+06	0.0034
L27	0.1	3e+19	10	1e+09	1e-11	1e-11	1e+07	0.6	1.9e+06	0.0019
L28	0.1	3e+21	10	1e+09	1e-12	1e-11	6e+07	0.6	1.5e+07	0.0150

**Extended Data Table 4 | Conditions and results of 2D seismomechanical numerical experiments for lithospheric deformation assisted by dyke propagation**

Run	$\eta_{\text{lith}}$ (Pa s)	Dyke length (km)	$P_{\text{lith}}$ (Pa)	$\lambda_{\text{melt(m)}}$	Strain rate (1/s)	$C$ (Pa)	$\phi$	$t_0$ (yr)	$t_a$ (yr)	$\tau_{\text{vol}}$ (Pa)	$\tau_{\text{eff}}$ (Pa)	$\lambda_{\text{melt}}$
D65	1e+23	1	1e+09	0.01	1e-12	1e+07	0.6	2.36	0.75	2.5e+8	1.7e+07	0.0113
D72	1e+23	1	1e+09	0.01	1e-12	1e+06	0.6	0.24	0.24	8.1e+7	7.1e+06	0.0101
D73	1e+23	1	1e+09	0.01	1e-12	1e+06	0.2	0.24	0.24	7.9e+7	5.9e+06	0.0247
D74	1e+23	1	1e+09	0.001	1e-12	1e+06	0.2	0.24	0.24	7.8e+7	5.9e+06	0.0243
D75	1e+23	2	1e+09	0.001	1e-12	1e+06	0.2	0.24	0.24	4.6e+7	4.1e+06	0.0154
D76	1e+23	2	1e+09	0.0001	1e-12	1e+06	0.2	0.24	0.24	4.7e+7	3.8e+06	0.0142
D78	1e+23	1	1e+09	0.01	1e-13	1e+07	0.6	23.6	7.44	2.8e+8	2.3e+07	0.0214
D79	1e+24	1	1e+09	0.01	1e-13	1e+07	0.6	23.6	7.48	2.9e+8	1.9e+07	0.0150
D88	1e+22	1	1e+09	0.01	1e-13	1e+07	0.6	23.6	7.48	2.7e+8	5.2e+07	0.0701
D92	1e+23	1	2e+09	0.01	1e-12	1e+07	0.6	2.36	2.32	5.3e+8	2.8e+07	0.0150
D93	1e+23	1	1e+09	0.001	1e-12	1e+06	0.6	0.24	0.24	8.0e+7	5.3e+06	0.0072
D99	1e+23	1	1e+09	0.001	1e-13	1e+06	0.6	2.36	0.24	4.8e+7	3.3e+06	0.0039
D100	1e+23	1	1e+09	0.001	1e-13	1e+06	0.2	2.36	0.24	4.6e+7	3.2e+06	0.0111
D201	1e+23	2	1e+09	0.0001	1e-12	1e+06	0.2	0.24	0.75	1.0e+8	1.9e+07	0.0885

# Widespread exploitation of the honeybee by early Neolithic farmers

Mélanie Roffet-Salque<sup>1</sup>, Martine Regert<sup>2</sup>, Richard P. Evershed<sup>1</sup>, Alan K. Outram<sup>3</sup>, Lucy J. E. Cramp<sup>1,4</sup>, Orestes Decavallas<sup>5,6</sup>, Julie Dunne<sup>1</sup>, Pascale Gerbault<sup>7,8</sup>, Simona Mileto<sup>1,9</sup>, Sigríð Mirabaud<sup>6†</sup>, Mirva Pääkkönen<sup>1,10</sup>, Jessica Smyth<sup>1,4</sup>, Lucija Šoberl<sup>1,11†</sup>, Helen L. Whelton<sup>1</sup>, Alfonso Alday-Ruiz<sup>12</sup>, Henrik Asplund<sup>10</sup>, Marta Bartkowiak<sup>13</sup>, Eva Bayer-Niemeier<sup>14</sup>, Lotfi Belhouchet<sup>15</sup>, Federico Bernardini<sup>16,17</sup>, Mihael Budja<sup>11</sup>, Gabriel Cooney<sup>18</sup>, Miriam Cubas<sup>19†</sup>, Ed M. Danaher<sup>20</sup>, Mariana Diniz<sup>21</sup>, László Domboróczki<sup>22</sup>, Cristina Fabbri<sup>23</sup>, Jesus E. González-Urquijo<sup>19</sup>, Jean Guilaine<sup>24</sup>, Slimane Hachi<sup>25</sup>, Barrie N. Hartwell<sup>26</sup>, Daniela Hofmann<sup>27</sup>, Isabel Hohle<sup>28</sup>, Juan J. Ibáñez<sup>29</sup>, Necmi Karul<sup>30</sup>, Farid Kherbouche<sup>25</sup>, Jacinta Kiely<sup>31</sup>, Kostas Kotsakis<sup>32</sup>, Friedrich Lueth<sup>33</sup>, James P. Mallory<sup>26</sup>, Claire Manen<sup>24</sup>, Arkadiusz Marciniak<sup>13</sup>, Brigitte Maurice-Chabard<sup>34</sup>, Martin A. Mc Gonigle<sup>35</sup>, Simone Mulazzani<sup>36,37</sup>, Mehmet Özdoğan<sup>30</sup>, Olga S. Perić<sup>38</sup>, Slaviša R. Perić<sup>38</sup>, Jörg Petrasch<sup>39</sup>, Anne-Marie Pétrequin<sup>40</sup>, Pierre Pétrequin<sup>40</sup>, Ulrike Poensgen<sup>41</sup>, C. Joshua Pollard<sup>42</sup>, François Poplin<sup>43</sup>, Giovanna Radi<sup>23</sup>, Peter Stadler<sup>44</sup>, Harald Stäuble<sup>45</sup>, Nenad Tasić<sup>46</sup>, Dushka Urem-Kotsou<sup>47</sup>, Jasna B. Vuković<sup>46</sup>, Fintan Walsh<sup>48</sup>, Alasdair Whittle<sup>49</sup>, Sabine Wolfram<sup>50</sup>, Lydia Zapata-Peña<sup>12‡</sup> & Jamel Zoughlami<sup>51</sup>

The pressures on honeybee (*Apis mellifera*) populations, resulting from threats by modern pesticides, parasites, predators and diseases, have raised awareness of the economic importance and critical role this insect plays in agricultural societies across the globe. However, the association of humans with *A. mellifera* predates post-industrial-revolution agriculture, as evidenced by the widespread presence of ancient Egyptian bee iconography dating to the Old Kingdom (approximately 2400 BC)<sup>1</sup>. There are also indications of Stone Age people harvesting bee products; for example, honey hunting is interpreted from rock art<sup>2</sup> in a prehistoric Holocene context and a beeswax find in a pre-agriculturalist site<sup>3</sup>. However, when and where the regular association of *A. mellifera* with agriculturalists emerged is unknown<sup>4</sup>. One of the major products of *A. mellifera* is beeswax, which is composed of a complex suite of lipids including *n*-alkanes, *n*-alkanoic acids and fatty acyl wax esters. The composition is highly constant as it is determined genetically

through the insect's biochemistry. Thus, the chemical 'fingerprint' of beeswax provides a reliable basis for detecting this commodity in organic residues preserved at archaeological sites, which we now use to trace the exploitation by humans of *A. mellifera* temporally and spatially. Here we present secure identifications of beeswax in lipid residues preserved in pottery vessels of Neolithic Old World farmers. The geographical range of bee product exploitation is traced in Neolithic Europe, the Near East and North Africa, providing the palaeoecological range of honeybees during prehistory. Temporally, we demonstrate that bee products were exploited continuously, and probably extensively in some regions, at least from the seventh millennium cal BC, likely fulfilling a variety of technological and cultural functions. The close association of *A. mellifera* with Neolithic farming communities dates to the early onset of agriculture and may provide evidence for the beginnings of a domestication process.

<sup>1</sup>Organic Geochemistry Unit, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK. <sup>2</sup>CEPAM – Cultures et Environnements. Préhistoire, Antiquité, Moyen Âge, UMR 7264, Université Nice Sophia Antipolis – CNRS, 06300 Nice, France. <sup>3</sup>Department of Archaeology, University of Exeter, Laver Building, North Park Road, Exeter, Devon EX4 4QE, UK. <sup>4</sup>Department of Archaeology and Anthropology, University of Bristol, 43 Woodland Road, Bristol BS8 1UU, UK. <sup>5</sup>Université Bordeaux Montaigne, 33607 Pessac, France. <sup>6</sup>Laboratoire du Centre de Recherche et de Restauration des Musées de France (C2RMF), UMR 171, Palais du Louvre, Porte des Lions, 14 Quai François Mitterrand, 75001 Paris, France. <sup>7</sup>Research Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK. <sup>8</sup>Department of Anthropology, University College London, London WC1H 0BW, UK. <sup>9</sup>Institut für Prähistorische Archäologie, Freie Universität Berlin, Altensteinstr. 15, Berlin 14195, Germany. <sup>10</sup>Department of Archaeology, University of Turku, 20014 Turun Yliopisto, Finland. <sup>11</sup>University of Ljubljana, Faculty of Arts, Department of Archaeology, Aškerčeva 2, box 580, 1000 Ljubljana, Slovenia. <sup>12</sup>Department of Geography, Prehistory and Archaeology, University of Basque Country (EHU-UPV), Francisco Tomás y Valiente s/n, 01006 Vitoria-Gasteiz, Spain. <sup>13</sup>Institute of Prehistory, Adam Mickiewicz University, Umultowska 89d, 61-614 Poznań, Poland. <sup>14</sup>Museum Quintana – Archäologie in Künzing, Partnermuseum der Archäologischen Staatssammlung München, Osterhofener Str. 2, 94550 Künzing, Germany. <sup>15</sup>Musée Archéologique de Sousse, Rue Marshall Tito, 4000 Sousse, Tunisia. <sup>16</sup>Centro Fermi, Museo Storico della Fisica e Centro di Studi e Ricerche Enrico Fermi, 00184 Rome, Italy. <sup>17</sup>Multidisciplinary Laboratory, The Abdus Salam International Centre for Theoretical Physics, 34151 Trieste, Italy. <sup>18</sup>UCD School of Archaeology, University College Dublin, Dublin 4, Ireland. <sup>19</sup>International Institute for Prehistoric Research of Cantabria, University of Cantabria, Avd de los Castros s/n, 39005 Santander, Spain. <sup>20</sup>Department of Archaeology, University College Galway, Galway, Ireland. <sup>21</sup>UNIARQ-Departamento de História, Faculdade de Letras de Lisboa, Universidade de Lisboa, 1600-214 Lisboa, Portugal. <sup>22</sup>István Dobó Castle Museum, Vár út 1, 3300 Eger, Hungary. <sup>23</sup>Dipartimento Civiltà e Forme del Sapere, Università di Pisa, Via Galvani 1, 56126 Pisa, Italy. <sup>24</sup>CNRS – UMR 5608 – TRACES, Maison de la recherche, Université Toulouse Jean Jaurès, 5 Allée Antonio Machado, 31058 Toulouse cedex 9, France. <sup>25</sup>CNRP/PAH, Centre National de Recherche Préhistorique, Anthropologique et Historique, Algiers, Algeria. <sup>26</sup>School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, Belfast BT7 1NN, UK. <sup>27</sup>Universität Hamburg, Archäologisches Institut, Edmund-Siemers-Allee 1, Flügel West, 20146 Hamburg, Germany. <sup>28</sup>a.r.t.e.s. Graduate School for the Humanities Cologne, Graduiertenschule der Philosophischen Fakultät, Aachener Str. 217, 50931 Cologne, Germany. <sup>29</sup>IMF-CSIC, Egiptiacas 15, 08001 Barcelona, Spain. <sup>30</sup>Istanbul University, Faculty of Letters, Department of Prehistory, 34434 Laleli Istanbul, Turkey. <sup>31</sup>Eachtra Archaeological Projects, Lickbeg, Clashmore, County Waterford, Ireland. <sup>32</sup>School of History and Archaeology, Faculty of Philosophy, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece. <sup>33</sup>German Archaeological Institute, Podbielskiallee 69-71, 14 195 Berlin, Germany. <sup>34</sup>Musée Rolin, 3 rue des Bains, 71400 Autun, France. <sup>35</sup>John Cronin & Associates, 28 Upper Main Street, Buncrana, County Donegal, Ireland. <sup>36</sup>Aix-Marseille Université, CNRS, Ministère de la Culture et de la Communication, UMR 7269 LAMPEA, LabexMed, 13284 Marseille, France. <sup>37</sup>Dipartimento di Biologia Ambientale, Università degli Studi di Roma La Sapienza, Rome 00185, Italy. <sup>38</sup>Institute of Archaeology Belgrade, Kneza Mihaila 35/4 11000 Belgrade, Serbia. <sup>39</sup>Eberhard-Karls-Universität Tübingen, Institut für Ur- und Frühgeschichte und Archäologie des Mittelalters – Abt. Jüngere Urgeschichte und Frühgeschichte – Schloß Hohentübingen, 72070 Tübingen, Germany. <sup>40</sup>Maison des Sciences de l'Homme et de l'Environnement C.N. Ledoux, CNRS & Université de Franche-Comté, 32 rue Mégevand, 25030 Besançon Cedex, France. <sup>41</sup>Kämpfstr. 20, 78315 Radolfzell, Germany. <sup>42</sup>Department of Archaeology, Faculty of Humanities, University of Southampton, Avenue Campus, Highfield, Southampton SO17 1BF, UK. <sup>43</sup>Muséum National d'Histoire Naturelle, 55 rue de Buffon, 75005 Paris, France. <sup>44</sup>Department of Pre- and Protohistory, University of Vienna, 1190 Vienna, Austria. <sup>45</sup>Landesamt für Archäologie, Zur Wetterwarte 7, 01109 Dresden, Germany. <sup>46</sup>Department of Archaeology, Faculty of Philosophy, Belgrade University, 18-20 Čika Ljubina Street, 11000 Belgrade, Serbia. <sup>47</sup>Department of History and Ethnology, Democritus University of Thrace, Komotini, Greece. <sup>48</sup>Irish Archaeological Consultancy, Unit G1, Network Enterprise Park, Kilcoole, County Wicklow, Ireland. <sup>49</sup>Department of Archaeology and Conservation, Cardiff University, John Percival Building, Colum Drive, Cardiff CF10 3EU, UK. <sup>50</sup>State Museum of Archaeology Chemnitz, Stefan-Heym-Platz 1, 09111 Chemnitz, Germany. <sup>51</sup>Institut National du Patrimoine de Tunis – Musée archéologique de Carthage, Carthage, Tunisia. †Present addresses: Département des restaurateurs, Institut National du Patrimoine, 124 rue Henri Barbusse, 93300 Aubervilliers, France (S.M.); Laboratório HERCULES, Universidade de Évora, Palácio do Vimioso, Largo Marquês de Marialva 8, 7000-809 Évora, Portugal (L.S.); BioArch-University of York, York YO10 5DD, UK (M.C.). ‡Deceased.



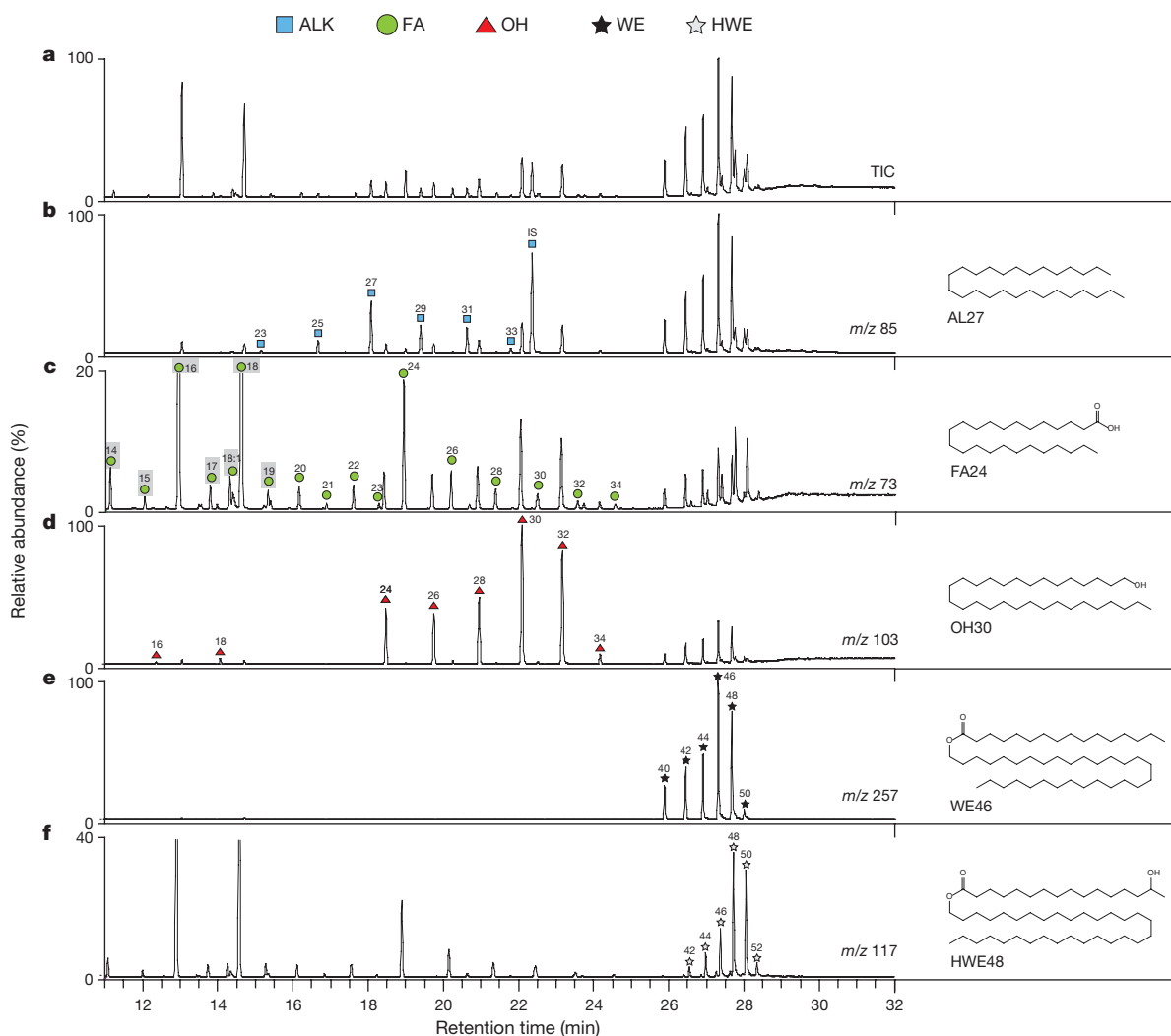
The honeybee holds a unique place in human culture. Notwithstanding its present-day economic importance, it has been revered over the millennia for the sheer beauty and complexity of the social organization within its colonies. For these reasons the honeybee is the most researched of the social insects, with its origin being regularly considered<sup>5</sup>. The last Ice Age would have had a major effect on the honeybee with the ice sheets restricting European populations to the northern Mediterranean hinterlands<sup>6</sup>. With the glacial retreat, the population would have subsequently expanded northwards. However, due to the lack of a Holocene fossil record<sup>7</sup>, the honeybee is ecologically invisible for most of the past 10,000 years.

Intriguingly, this is the period during which Neolithic agriculture emerged and spread out of southeastern Anatolia and the Levant, with some human population movement into ecological zones also conducive to the honeybee. Indeed, progressive woodland clearances by pioneer prehistoric farmers may have opened up forests, favouring light-demanding shrubs, herbs and fruit trees (for example, *Rosaceae*)<sup>8</sup>. Whether this would have exerted negative or positive effects on honeybee populations is unknown<sup>8,9</sup>. Given the latter, an opportunity exists to investigate the presence and early exploitation of the honeybee by prehistoric farming communities through the cultural materials recovered

from Neolithic sites, namely their recently invented pottery vessels, and in doing so, to assess the palaeoecological range of the honeybee in the Holocene.

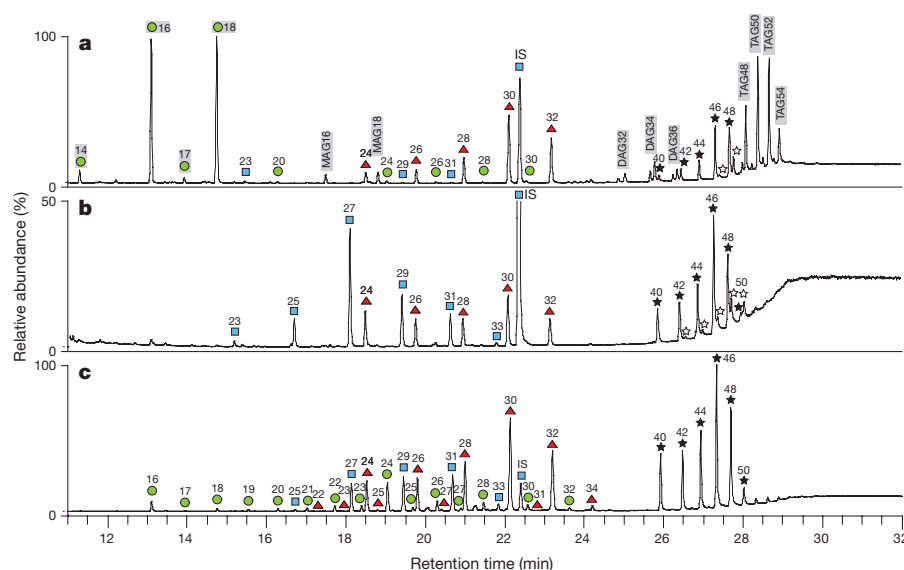
Although the most obvious reason for exploiting the honeybee would be for honey, a rare source of sweetener for prehistoric people, beeswax would likely have been an equally important material for various technological, ritual, cosmetic and medicinal applications<sup>10</sup>. Indeed, beeswax has been regularly detected in later archaeological and historic periods in lipid extracts from the fabric of unglazed pottery vessels<sup>11</sup> where it is assumed to be a residue of honey use in cooking, or from the use of vessels for processing wax combs<sup>12–14</sup>, with beeswax being absorbed through repeated contacts. Beeswax has also been detected as a fuel in lamps and in larger vessels used as proto-beehives, for example Roman Greece (second century BC to fourth century AD)<sup>15,16</sup> and applied as a post-firing treatment to waterproof vessels<sup>17</sup>.

The detection of beeswax in archaeological and historic contexts rests on its complex chemistry providing a unique and relatively recalcitrant chemical signature. Fresh beeswax comprises a complex mixture of aliphatic compounds consisting of series of homologues differing in chain-length by two methylene groups<sup>18</sup>. Medium-chain *n*-alkanes range from C<sub>23</sub> to C<sub>31</sub> (with C<sub>27</sub> dominating in *A. mellifera*),



**Figure 1 | High-temperature gas chromatography/mass spectrometry chromatograms of total lipid extract of a sherd from Çayönü Tepesi (6500–6000 cal BC) containing beeswax.** a–f, Partial total ion current chromatogram (a) and mass chromatograms (b–f) displaying ion masses of characteristic fragments from the main compound classes comprising the extract ( $m/z$  85, 73, 103, 257 and 117, respectively) with the molecular structure of the most abundant component for each compound class.

Squares, *n*-alkanes (ALK); circles, *n*-alkanoic acids (fatty acids, FA); triangles, *n*-alkanols (OH); black asterisks, fatty acyl monoesters (WE); grey asterisks, hydroxyl fatty acyl monoesters (HWE); IS, internal standard (*n*-tetratriacontane); number *n* and *n*:*i*, acyl carbon number with zero or *i* degrees of unsaturations. Compounds shown with a grey background are interpreted as originating from mammalian animal fats.



**Figure 2 | Partial gas chromatograms of total lipid extracts from Neolithic sherds from each geographical region. a, Mağura (5500–5200 cal BC). b, Niederhummel (5360–5220 cal BC). c, Gueldaman (fifth**

millennium BC). **a** is interpreted as mixture of animal fats and beeswax; **b** and **c** as pure beeswax. MAG, monoacylglycerols; DAG, diacylglycerols, TAG, triacylglycerols. Other peak attributions as in Fig. 1.

and *n*-alkanoic acids from C<sub>20</sub> to C<sub>36</sub>. Monoesters comprise predominantly alkyl palmitates (C<sub>38</sub> to C<sub>52</sub>), with characteristic hydroxy monoesters comprising long-chain alcohols (C<sub>24</sub> to C<sub>38</sub>) esterified mainly to hydroxypalmitic acid, ranging between C<sub>40</sub> and C<sub>54</sub> (ref. 18). The hydrophobic nature of beeswax makes it relatively resistant to degradation. Hence, if protected from extensive microbial attack and/or exposure to high temperatures during anthropogenic manipulation, the aforementioned chemical characteristics can be used in assessing its presence<sup>10,19</sup> (Figs 1 and 2).

Adopting this lipid biomarker approach, we now explore the association of the honeybee with the spread of early Old World farmers based on lipid residue analyses of more than 6,400 pottery vessels (Supplementary Information sections 1 and 2). Combining our new findings with published occurrences of beeswax in prehistoric pottery allows the association between honeybees and early farmers to be mapped spatially and temporally through prehistory (Figs 3 and 4).

The oldest evidence for beeswax comes from Neolithic sites in Anatolia dating from the seventh millennium cal BC, as these sites are the locations of the oldest pottery vessels in Europe and Eurasia. Most of the assemblages investigated comprised globular or bowl shape 'cooking' vessels, an interpretation supported by the finding of ruminant and porcine animal fats in significant numbers of vessels. No beeswax residues were detected during the intensive investigations of >380 vessels from the Levant, although only 34 residues were detected<sup>20</sup>. Moving into eastern Anatolia, the site of Çayönü Tepesi revealed two beeswax residues from 83 vessels from the seventh millennium including an exceptionally well-preserved residue containing all the biomarkers of beeswax (Fig. 1b–f). The free *n*-alkanols, dominated by C<sub>30</sub> and C<sub>32</sub> homologues, do not occur in fresh beeswax but are a feature of aged wax, due to hydrolysis of the wax esters. The high abundance of C<sub>18:0</sub> fatty acid suggests mixing with mammalian animal fat, the latter being common in other sherds in the assemblage<sup>20</sup>. The second sherd from this site contained a lower concentration of beeswax but all the biomarkers were clearly evident. These two residues establish the easterly limit of the beeswax detected in this investigation and provide the oldest unequivocal evidence, to our knowledge, of honeybee exploitation by early Neolithic farmers.

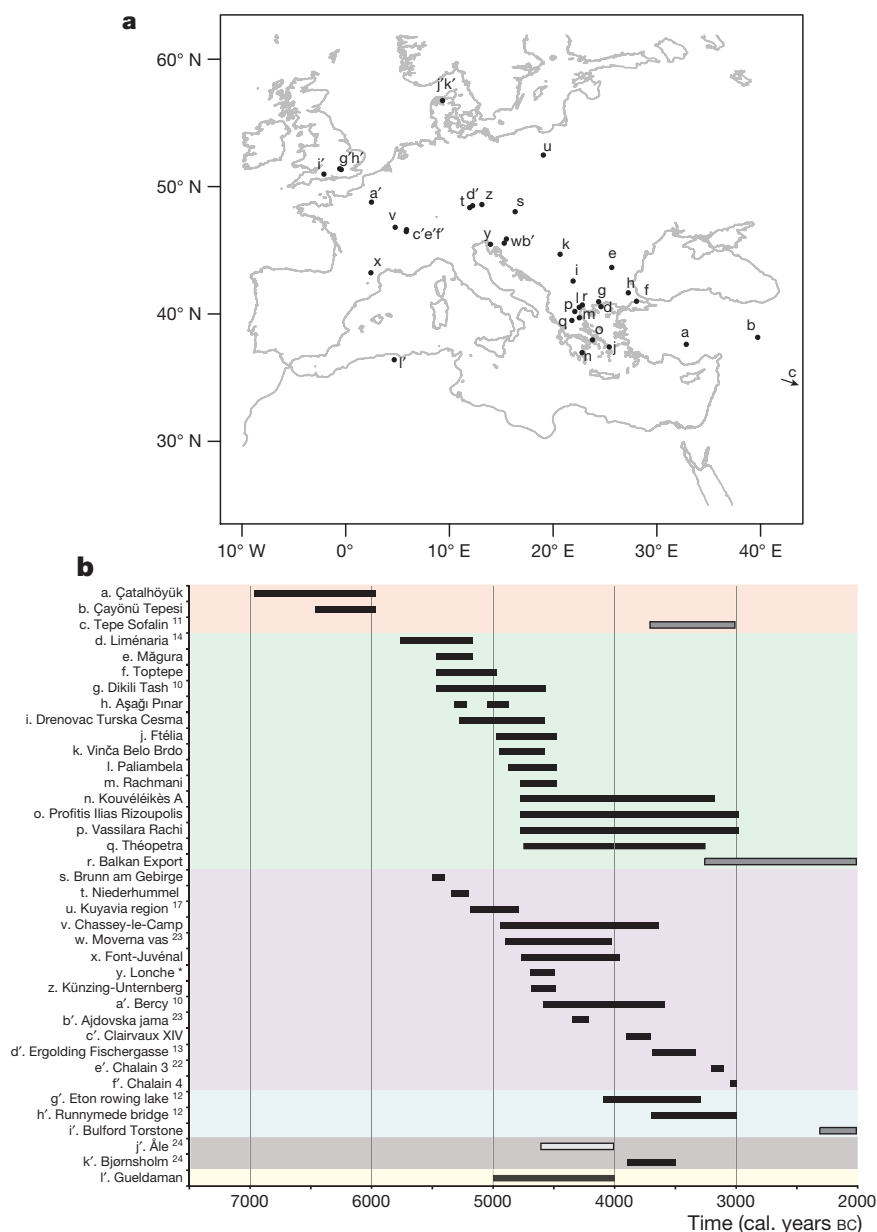
In central Anatolia, extensive investigations of organic residues in 650 vessels, mainly from the site of Çatalhöyük, revealed abundant animal fat residues. Only one residue showed tentative evidence for beeswax based on wax esters, dominated by C<sub>46</sub> and C<sub>48</sub> homologues; however, the *n*-alkanols do not exhibit the familiar distribution.

*n*-Alkanes were detectable but the distribution is skewed towards the higher homologues compared to that expected in fresh beeswax, although such distributional changes are frequently seen in historical and archaeological beeswax, assumed to arise by sublimation during ageing or heat treatment<sup>10</sup>. The tentative identification of this very early beeswax residue at Çatalhöyük is supported by the discovery of a striking depiction of a honeycomb-like pattern painted on a wall at the site<sup>21</sup>.

Analyses of approximately 570 cooking vessels from northwestern Anatolia revealed 72 lipid residues of which 4 were identified as containing beeswax, from Aşağı Pinar and Toptepe, dating to 5500–5000 cal BC. Although the overall purity of the beeswax (two were mixed with ruminant fat) and lipid concentrations (20 to 220 µg per gram of sherd) were quite variable, the distributions were unmistakable. One of the beeswax finds from Toptepe is well preserved, albeit with ageing evident from the hydrolytically released free *n*-alkanols and slight distortions of the various homologous series, through loss of lower homologues.

The most abundant evidence for honeybee exploitation by early farmers was seen in the rest of the Balkan Peninsula. The full range of beeswax biomarkers was identified in sherds from bowls, pans and sieves from the Late Neolithic sites of Paliambela, Greece (4900–4500 cal BC), Mağura, Romania (Fig. 2a; 5500–5200 cal BC) and Drenovac Turska Česma, Serbia (5300–4700/4600 cal BC). A large number of beeswax residues were found in Neolithic potsherds (11 residues out of 81 sherds analysed) from Attica, the Peloponnese and the Cyclades (Aegean Islands), dating between 5800 and 3000 cal BC, firmly establishing the long tradition of bee exploitation in this region. Overall, the incidence of beeswax residues is highest in the Balkan Peninsula, where of the 1,915 Neolithic sherds analysed, 473 yielded lipid residues, of which 5.5% contained beeswax.

In Central Europe, pure beeswax was recovered from potsherds from *Linearbandkeramik* (LBK) sites occupied by the earliest farmers of Austria and Germany (oldest LBK) including the sites of Brunn am Gebirge (5500–5400 cal BC) and Niederhummel (5360–5220 cal BC), pushing back the date for bee exploitation in this region by approximately 1,500 years<sup>13</sup> (Fig. 2b). Beeswax was also detected in late sixth millennium LBK sites of Ludwinowo 7 and Wolica Nowa, Poland<sup>17</sup>. In France, the exploitation of bee products is evident during the second half of the fifth millennium at *Chasséen* sites (Font-Juvénal, Chasseyle-Camp and Bercy<sup>10</sup>) and fourth millennium at the Lake Village sites of Clairvaux-les-Lacs (3900 to 3700 BC) and Chalais 3 (ref. 22) and 4 (3200 to 3100 BC and 3040 to 2990 BC). High incidence of beeswax (approximately 15% of the detectable residues) was identified in



**Figure 3 | Geographical distribution of prehistoric sites in the date range 7500 and 2000 cal BC yielding beeswax residues. a, Locations of archaeological sites. b, Chronology of beeswax use in the Near East, the Balkan Peninsula, mainland Europe, Scandinavia, the UK and northern**

Africa. Neolithic finds in black, pre-Neolithic (hunter-gatherer contexts) in light grey and Bronze Age in dark grey. \* Dental filling re-examined after ref. 31.

fifth millennium sherds from two Slovenian sites (Ajdovska jama and Moverna vas)<sup>23</sup>.

Around 130 sherds have so far been analysed from the Iberian Peninsula. However, no beeswax residues have yet been detected, although the overall preservation of organic residues was poor. Further investigations will likely reveal examples of beeswax in Neolithic pottery from this region.

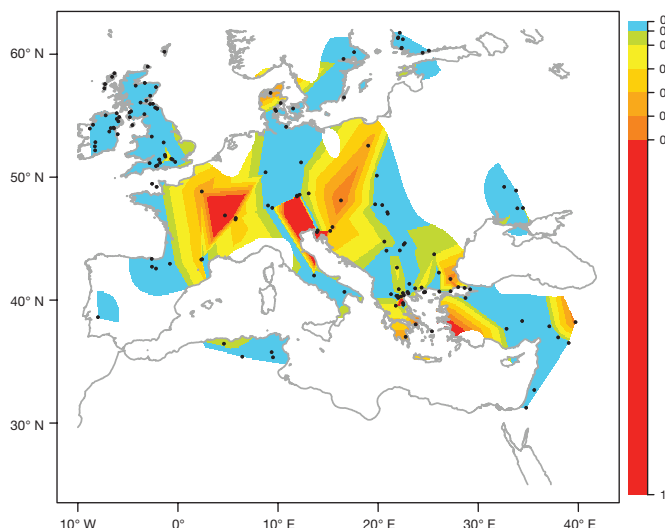
The northerly limit of bee exploitation in northern Europe appears to be Denmark with two beeswax finds in late Mesolithic and Neolithic contexts<sup>24</sup>. Around 5° to the south in southern Britain, beeswax is evident in 7 vessels amongst the approximately 670 Neolithic vessels analysed. These findings clearly counter any arguments for a late introduction of the honeybee into the British Isles<sup>8,25</sup>. Interestingly, however, investigations of nearly 1,200 Mesolithic and Neolithic vessels from Ireland, Scotland and Fennoscandia<sup>26,27</sup> have failed to reveal any conclusive evidence of beeswax (Supplementary Information section 3). Given that organic residue preservation in these regions is excellent,

the lack of beeswax would seem to establish the ecological limit of *A. mellifera* at that time. Similar arguments are likely to account for the absence of beeswax residues from >350 prehistoric pottery vessels from the Eurasian Steppe<sup>28</sup>.

Finally, we report the first evidence for bee exploitation by Neolithic pastoralists in North Africa. The analysis of 71 sherds from the Algerian site of Gueldaman revealed a single well-preserved beeswax residue (fifth millennium BC). The preservation is again exceptional with *n*-alkanes, *n*-fatty acids and fatty acyl wax ester distributions providing an unequivocal identification of beeswax. The presence of free long-chain *n*-alkanols and lack of hydroxy fatty acid wax esters are indicative of diagenesis and/or use-related alteration. However, the overall distribution indicates the wax residue derives from *A. mellifera* (Fig. 2c).

In conclusion, the approximately 50 new finds of beeswax residues considered above provide evidence for the widespread exploitation of the honeybee by the early agriculturalists and pastoralists of the Near East, Europe and North Africa dating back nearly 9,000 years





**Figure 4 | Regional distribution of beeswax residues in potsherd lipid extracts.** Interpolated map of Old World beeswax occurrences (proportion of beeswax residues per number of residues in pottery sherds, in percentages) during the Neolithic (including the Mesolithic sites available). Colours and colour key show the proportions of beeswax residues estimated by surface interpolation, where collection locations are represented by dots ( $n = 154$ ).

(Fig. 3). In all these regions the new data have either provided the first evidence of honeybee exploitation in a region, as in North Africa, or pushed the chronology of human–honeybee association to substantially earlier dates, as in Anatolia and Central Europe (Fig. 3b). The lack of evidence for beeswax use at Neolithic sites north of the 57th parallel North may suggest an ecological limit to the natural occurrence of honeybees. Indeed, harsh high-latitude conditions, even with temperatures warmer than today<sup>29</sup>, would affect the foraging capabilities of honeybees<sup>30</sup>. Critically, in the absence of a Holocene fossil record for *A. mellifera*<sup>7</sup> these findings provide the first ancient biomolecule-based palaeoecological map of the distribution of an economically and culturally important animal (Fig. 4).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 23 April; accepted 29 September 2015.**

- Crane, E. *The World History of Beekeeping and Honey Hunting* (Duckworth, 1999).
- Dams, M. & Dams, L. R. Spanish art rock depicting honey gathering during the Mesolithic. *Nature* **268**, 228–230 (1977).
- d'Errico, F. *et al.* Early evidence of San material culture represented by organic artifacts from Border Cave, South Africa. *Proc. Natl Acad. Sci. USA* **109**, 13214–13219 (2012).
- Larson, G. & Fuller, D. Q. The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.* **45**, 115–136 (2014).
- Wallberg, A. *et al.* A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genet.* **46**, 1081–1088 (2014).
- Ruttner, F. *Biogeography and Taxonomy of Honeybees* (Springer, 1987).
- Buckland, P. I. & Buckland, P. C. in *Versions: BugsCEP v7.63; Bugsdata v8.01; BugsMCR v2.02; BugStats v1.22* (2006).
- Limbrey, S. in *Archaeological Aspects of Woodland Ecology* (eds Bell, M. & Limbrey, S.) 279–286 (British Archaeological Reports, 1982).
- Wilson, E. O. *The Insect Societies* (Belknap Press of Harvard Univ. Press, 1971).
- Regert, M., Colinart, S., Degrand, L. & Decavallas, O. Chemical alteration and use of beeswax through time: accelerated ageing tests and analysis of archaeological samples from various environmental contexts. *Archaeometry* **43**, 549–569 (2001).

- Mayyas, A. S., Al-Qudah, M. A., Douglas, K. A. & Al-Ajlouny, F. K. Beeswax preserved in archaeological ceramics: function and use. *Annals of Faculty of Arts. Ain Shams University* **40**, 343–371 (2012).
- Copley, M. S. *et al.* Dairying in antiquity. III. Evidence from absorbed lipid residues dating to the British Neolithic. *J. Archaeol. Sci.* **32**, 523–546 (2005).
- Heron, C., Nemcek, N., Bonfield, K. M., Dixon, D. & Ottaway, B. S. The chemistry of Neolithic beeswax. *Naturwissenschaften* **81**, 266–269 (1994).
- Decavallas, O. in *Cooking up the Past: Food and Culinary Practices in the Neolithic and Bronze Age Aegean* (eds Mee, C. & Renard, J.) 148–157 (Oxford Books Limited, 2007).
- Evershed, R. P., Vaughan, S. J., Dudd, S. N. & Soles, J. S. Fuel for thought? Beeswax in lamps and conical cups from Late Minoan Crete. *Antiquity* **71**, 979–985 (1997).
- Evershed, R. P., Dudd, S. N., Anderson-Stojanovic, V. R. & Gebhard, E. R. New chemical evidence for the use of combed ware pottery vessels as beehives in ancient Greece. *J. Archaeol. Sci.* **30**, 1–12 (2003).
- Salque, M. *et al.* Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature* **493**, 522–525 (2013).
- Aichholz, R. & Lorbeer, E. Investigation of combwax of honeybees with high-temperature gas chromatography and high-temperature gas chromatography–chemical ionization mass spectrometry: I. high-temperature gas chromatography. *J. Chromatogr. A* **855**, 601–615 (1999).
- Garnier, N., Cren-Olivé, C., Rolando, C. & Regert, M. Characterization of archaeological beeswax by electron ionization and electrospray ionization mass spectrometry. *Anal. Chem.* **74**, 4868–4877 (2002).
- Evershed, R. P. *et al.* Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* **455**, 528–531 (2008).
- Mellaart, J. Excavations at Çatal Hüyük, 1962: second preliminary report. *Anatolian Studies* **13**, 43–103 (1963).
- Regert, M., Dudd, S. N., Van Bergen, P. F., Pétrequin, P. & Evershed, R. P. Investigations of solvent extractable lipids and insoluble polymeric components: organic residues in Neolithic ceramic vessels from Chalais (Jura, France). *British Archaeological Reports* **939**, 78–90 (2001).
- Soberl, L. *et al.* Neolithic and Eneolithic activities inferred from organic residue analysis of pottery from Mala Triglavca, Moverna vas and Ajdovska jama, Slovenia. *Documenta Praehistorica* **41**, 149–179 (2014).
- Heron, C., Craig, O. E., Forster, M. & Stern, B. & Andersen, S. H. in *Shell Middens in Atlantic Europe* (eds Milner, N., Craig, O. E. & Bailey, G. N.) Ch. 7, 78–85 (Oxbow Books, 2007).
- Carreck, N. Are honey bees (*Apis mellifera* L.) native to the British Isles? *J. Apic. Res.* **47**, 318–322 (2008).
- Cramp, L. J. E. *et al.* Immediate replacement of fishing with dairying by the earliest farmers of the northeast Atlantic archipelagos. *Proc. R. Soc. Lond. B* **281**, 20132372 (2014).
- Cramp, L. J. E. *et al.* Neolithic dairy farming at the extreme of agriculture in northern Europe. *Proc. R. Soc. Lond. B* **281**, 20140819 (2014).
- Outram, A. K. *et al.* The earliest horse harnessing and milking. *Science* **323**, 1332–1335 (2009).
- Bartlein, P. J. *et al.* Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Clim. Dyn.* **37**, 775–802 (2011).
- Abou-Shaara, H. F. The foraging behaviour of honey bees, *Apis mellifera*: a review. *Vet. Med. (Praha)* **59**, 1–10 (2014).
- Bernardini, F. Beeswax as dental filling on a Neolithic human tooth. *PLoS ONE* **7**, e44904 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the UK Natural Environment Research Council for partial funding of the mass spectrometry facilities at Bristol (contract no. R8/H10/63; <http://www.lsmf.co.uk>) and English Heritage, European Research Council, Leverhulme Trust, Ministère de la Culture et de la Communication, Ministère de l'Enseignement Supérieur et de la Recherche (ACI Jeunes Chercheurs), Natural Environment Research Council, Région PACA, Royal Society and Wellcome Trust for funding.

**Author Contributions** M.R.-S., M.R., R.P.E. and A.K.O. conceived and planned the project about beeswax in prehistory. M.R.-S., M.R. and R.P.E. wrote the paper. M.R.-S., M.R., L.J.E.C., O.D., J.D., S.Mil., S.Mir., M.P., J.S., L.S., H.L.W., M.Bart. and D.U.-K. undertook planning of regional lipid residue analyses projects, sampling, analytical work and data analysis. P.G. created Figure 4 and Supplementary Information section 3. All other authors either directed excavations or provided expertise in relation to pottery collections and essential insights into the study region and sites.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. R code is available upon request to P.G. ([p.gerbault@ucl.ac.uk](mailto:p.gerbault@ucl.ac.uk)). Correspondence and requests for materials should be addressed to M.R.-S. ([melanie.salque@bristol.ac.uk](mailto:melanie.salque@bristol.ac.uk)), M.R. ([martine.regert@cepam.cnrs.fr](mailto:martine.regert@cepam.cnrs.fr)) or R.P.E. ([r.p.evershed@bristol.ac.uk](mailto:r.p.evershed@bristol.ac.uk)).

## METHODS

**Lipid residue analyses.** All solvents used were HPLC grade (Rathburn) and the reagents were analytical grade (typically >98% of purity).

A sub-sample (1 to 3 g) from archaeological potsherds was cleaned with a modelling drill to remove any exogenous lipids (from the soil and handling) and crushed with a solvent-washed mortar and pestle. An internal standard (*n*-tetratriacontane, typically 20 µg) was added to the powdered sherd to enable the quantification of lipid extract. Ground samples of sherds were extracted with CHCl<sub>3</sub>/MeOH (2:1 (vol/vol), 2 × 10 ml) using ultrasonication. Both supernatants were combined and the solvent was removed under a gentle stream of nitrogen at 40 °C. Aliquots of the total lipid extract (TLE) were treated with 40 µl of *N,O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) containing 1% trimethylchlorosilane (Sigma Aldrich) for 1 h at 70 °C and the BSTFA in excess evaporated under a gentle stream of nitrogen. The trimethylsilylated TLE was diluted in hexane (typically 50 to 150 µl) and submitted to analysis by high-temperature gas chromatography (HTGC) and high-temperature gas chromatography-mass spectrometry (HTGC/MS) to identify the major compounds present.

All TLEs were initially screened in a Hewlett-Packard 5890 Series II gas chromatograph equipped with a fused-silica capillary column (15 m × 0.32 mm) coated with dimethyl polysiloxane stationary phase (DB-1HT; film thickness, 0.1 µm; Agilent Technologies). Derivatized extracts (1.0 µl) were injected on-column using a cool on-column inlet in track oven mode. The temperature was held isothermally for 2 min at 5 °C and then increased at a rate of 10 °C min<sup>-1</sup> and held at 350 °C for 10 min. The flame ionization detector (FID) was set at a temperature of 350 °C. Helium was used as a carrier gas and maintained at a constant flow of 4.6 ml min<sup>-1</sup>. Data acquisition and processing were carried out using the HP Chemstation software (Rev. B.03.02 (341), Agilent Technologies).

HTGC/MS analyses of trimethylsilylated aliquots were performed using a Thermo Scientific Trace 1300 gas chromatograph coupled with an ISQ single quadrupole mass spectrometer. Diluted samples were introduced using a PTV injector in split mode (split flow of 30 ml min<sup>-1</sup>, split ratio of 6.0) onto a 0.53 mm fused silica pre-column connected to a 15 m × 0.32 mm i.d. fused-silica capillary column coated with dimethyl polysiloxane stationary phase (Rxi-1HT; film thickness, 0.1 µm; Restek). The initial injection port temperature was 50 °C with an evaporation phase of 0.05 min, followed by a transfer phase from 50 °C to 380 °C at 0.2 °C min<sup>-1</sup>. The oven temperature was held isothermally for 2 min at

50 °C, increased at a rate of 10 °C min<sup>-1</sup> to 280 °C, then at a rate of 25 °C min<sup>-1</sup> to 380 °C and finally held at 380 °C for 5 min. Helium was used as a carrier gas and maintained at a constant flow 5 ml min<sup>-1</sup>. The mass spectrometer was operated in the electron ionization (EI) mode (70 eV) with a GC interface temperature of 380 °C and a source temperature of 340 °C. The emission current was 50 µA and the mass spectrometry set to acquire in the range of *m/z* 50–950 Daltons at two scans per second. Data acquisition and processing were carried out using the Thermo XCalibur software (version 3.0.63). Peaks were identified on the basis of their mass spectra, gas chromatography (GC) retention times, by comparison with the NIST mass spectral library (version 2.0) and by comparison with modern beeswax (from the Loire department, France).

**Construction of Fig. 4.** The total number of archaeological sites investigated is 166, but only 154 of these fell within the geographical area of interest (longitude –10° to 42° and from latitude 25° to 62°, see Supplementary Information section 1). To estimate the distribution of beeswax residues in continuous space from irregularly spaced data, linear interpolation was performed in the triangles bounded by data points<sup>32,33</sup>. The output grid was made of 530 × 380 points evenly spaced over the range of latitude and longitude. No extrapolation was being used. Kriging was used to narrow the interpolation values to locations around data points (and not show interpolation values where there is no data). Kriging allows to obtain weights of the prediction locations based on the distance between data points, with lower variance where data points are and higher variance where there is no data. Interpolation, kriging and plotting were all performed in R version 2.15.1 (ref. 34). Interpolation was performed using the function 'interp' from the package 'akima' (CRAN repository, <http://cran.r-project.org/web/packages/akima/akima.pdf>). Kriging was performed using the function 'krige.conv' from the package 'geoR' (CRAN repository, <http://cran.r-project.org/web/packages/geoR/geoR.pdf>, further information on the package 'geoR' can be found at <http://www.leg.ufpr.br/geoR>). R code available upon request to P.G.

32. Akima, H. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Trans. Math. Softw.* **4**, 148–159 (1978).
33. Akima, H. Algorithm 761: scattered-data surface fitting that has the accuracy of a cubic polynomial. *ACM Trans. Math. Softw.* **22**, 362–371 (1996).
34. The R Core Team. R: a Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, 2012).

# Deep-time evolution of regeneration and preaxial polarity in tetrapod limb development

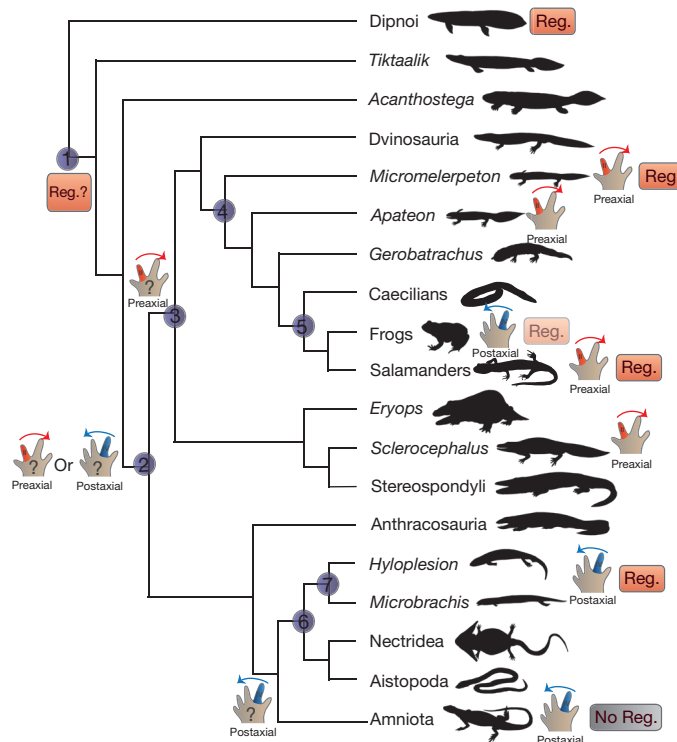
Nadia B. Fröbisch<sup>1</sup>, Constanze Bickelmann<sup>1</sup>, Jennifer C. Olori<sup>2</sup> & Florian Witzmann<sup>1,3</sup>

Among extant tetrapods, salamanders are unique in showing a reversed preaxial polarity in patterning of the skeletal elements of the limbs, and in displaying the highest capacity for regeneration, including full limb and tail regeneration. These features are particularly striking as tetrapod limb development has otherwise been shown to be a highly conserved process<sup>1,2</sup>. It remains elusive whether the capacity to regenerate limbs in salamanders is mechanistically and evolutionarily linked to the aberrant pattern of limb development; both are features classically regarded as unique to urodeles<sup>3</sup>. New molecular data suggest that salamander-specific orphan genes play a central role in limb regeneration and may also be involved in the preaxial patterning during limb development<sup>4,5</sup>. Here we show that preaxial polarity in limb development was present in various groups of temnospondyl amphibians of the Carboniferous and Permian periods, including the dissorophoids *Apateon* and *Micromelerpeton*, as well as the stereospondylomorph *Sclerocephalus*. Limb regeneration has also been reported in *Micromelerpeton*<sup>6</sup>, demonstrating that both features were already present together in antecedents of modern salamanders 290 million years ago. Furthermore, data from lepospondyl ‘microsaurs’ on the amniote stem indicate that these taxa may have shown some capacity for limb regeneration and were capable of tail regeneration<sup>7</sup>, including re-patterning of the caudal vertebral column that is otherwise only seen in salamander tail regeneration. The data from fossils suggest that salamander-like regeneration is an ancient feature of tetrapods that was subsequently lost at least once in the lineage leading to amniotes. Salamanders are the only modern tetrapods that retained regenerative capacities as well as preaxial polarity in limb development.

Research on tetrapod limb development is a prime example of how the integration of palaeontological and developmental data can provide novel insights into the evolution of a crucial organ system<sup>1</sup>. In recent decades, research has revealed many aspects of the molecular mechanisms involved in limb development, and has shown that tetrapod limb development is a remarkably conserved process<sup>1,2</sup>. It is therefore even more surprising that salamanders are an exception to this overall conserved pattern of limb development, characterized by preaxial rather than postaxial or central polarity in the development of the zeugopodial elements and digits<sup>3,8,9</sup>. Moreover, salamanders show the highest regenerative capacity among tetrapods, including full limb and tail regeneration throughout the lifetime of an individual<sup>10</sup>. Both features have classically been considered highly derived for salamanders<sup>9,11</sup>, yet surprisingly the question of whether the remarkable regenerative capacities of salamander limbs and the aberrant pattern of limb development are evolutionarily and/or mechanistically linked, has never been investigated, to our knowledge<sup>3,4</sup>. Here we present new evidence from the fossil record that provides a novel deep-time perspective on the evolution of preaxial polarity in limb development and salamander-like regenerative capacities among early tetrapods.

The phylogeny of Tetrapoda comprises basal stem taxa and two major lineages: first the Temnospondyli (Fig. 1, node 3), a diverse

group of Palaeozoic and Mesozoic anamniotes that includes the clade Dissorophoidea (Fig. 1, node 4), which is supported by most recent studies as having given rise to all modern amphibians or at least batrachians (frogs and salamanders)<sup>12,13</sup>; second, the lineage comprising Lepospondyli (Fig. 1, node 6), a clade of mostly small-bodied Palaeozoic anamniotes, and Amniota<sup>13</sup>. Some lepospondyl taxa also are proposed to be lissamphibian predecessors in an alternative hypothesis of amphibian origins<sup>14</sup>. A number of temnospondyl taxa have an excellent fossil record that has provided detailed insights into their ontogenetic development and life history patterns<sup>15</sup>. Particularly, the fossil record of the Early Permian Branchiosauridae is distinguished by a well resolved ontogenetic series, which allowed for the reconstruction of the ossification sequence in the limbs<sup>16</sup>. Ossification patterns in



**Figure 1 | Phylogenetic tree depicting major groups of tetrapods.**

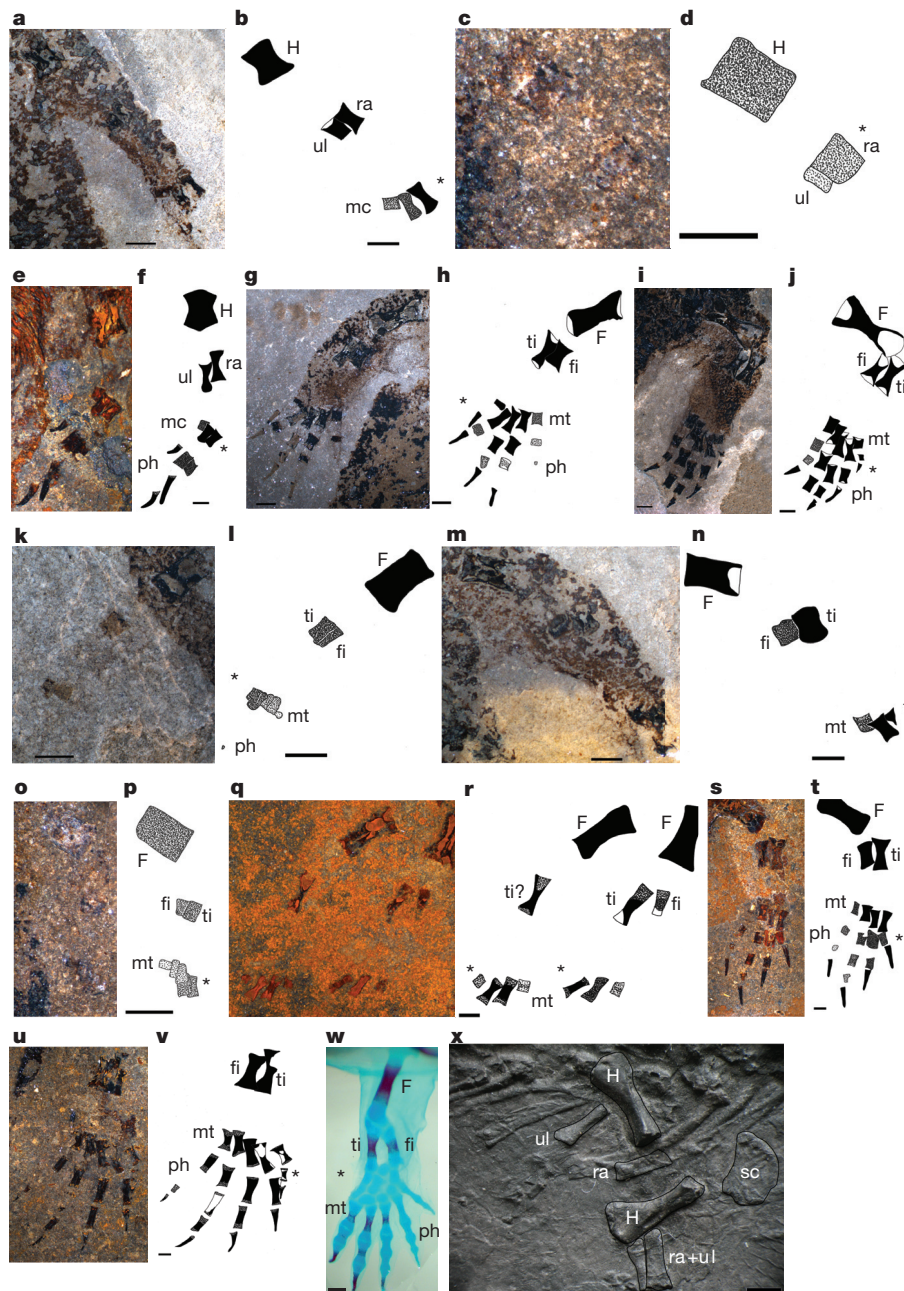
(1) Sarcopterygii; (2) crown-group tetrapods; (3) Temnospondyli; (4) Dissorophoidea; (5) Lissamphibia; (6) Lepospondyli; and (7) Microsauria. Stylized right hands indicate taxa for which pre- or postaxial polarity in limb development has been reported. Neither pre- nor postaxial polarity can be reconstructed as the plesiomorphic state for tetrapods (node 2). ‘Reg.’ indicates reported salamander-like regeneration; Reg. is depicted as faded for frogs, because they have regenerative capacities in limbs and tails only in tadpole stages until metamorphic climax is reached; ‘No Reg.’ indicates that regenerative capacity is absent; for taxa without a square, no data on regenerative capacity are currently available.

<sup>1</sup>Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Invalidenstrasse 43, 10115 Berlin, Germany. <sup>2</sup>Department of Biological Sciences, State University of New York at Oswego, 30 Centennial Drive, Oswego, New York 13126, USA. <sup>3</sup>Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA.



the limbs mirror general patterns of polarity of earlier events in skeletogenesis in extant and extinct tetrapods<sup>17</sup> and the ossification sequence in the branchiosaurid *Apateon* revealed that it indeed had preaxial polarity in limb development, previously exclusively attributed to salamanders<sup>16</sup>. However, it has remained uncertain whether this bears a phylogenetic signal or represents a homoplastic feature shared by branchiosaurids and salamanders owing to a similar larval ecology, as it was suggested that the early formation of the first and second digits of salamanders may be a larval adaptation<sup>9,18</sup>. The two temnospondyl taxa *Micromelerpeton* and *Sclerocephalus* derive from the same fossil lakes (*Lagerstätten*) in southwestern Germany as the branchiosaurid *Apateon*, and although the preserved ontogenetic sequences of these taxa are not as extensive as for *Apateon*, they provide detailed insights into the developmental trajectory of these species<sup>19,20</sup> (see Supplementary Information). *Micromelerpeton* represents the most basal member of the Dissorophoidea. Here we show for the first time that, as in branchiosaurids, fossils of successive ontogenetic larval stages of *Micromelerpeton* reveal a delayed ossification of the postaxial

zeugopodial element, as well as a clear gradation in the degree of ossification and differentiation of the metapodial and phalangeal elements from pre- to postaxial (Fig. 2a–o). This demonstrates that preaxial polarity in limb development was not only a feature of derived Dissorophoidea, but instead was already established at the base of Dissorophoidea. More surprisingly, new data in support of preaxial polarity in limb development were also found in *Sclerocephalus*, a stereospondylomorph temnospondyl and a much more distant relative of modern amphibians (Fig. 2p–w). While the preserved ontogenetic trajectory of *Sclerocephalus* is considerably less comprehensive than for *Apateon* and *Micromelerpeton*, a few very well-preserved larvae provide insights into patterns of limb ossification and indicate an advanced differentiation of preaxial zeugopodial and autopodial elements. Together, these new data on temnospondyl limb ossification patterns suggest an early evolution of preaxial polarity in limb development at the base of Temnospondyli or even at the base of Tetrapoda, as has been tentatively suggested previously<sup>3,16</sup> (Fig. 1). Notably, while it remains unknown whether *Sclerocephalus* and



**Figure 2 | Ossification patterns in the limbs of *Micromelerpeton* and *Sclerocephalus*.**

*Micromelerpeton* forelimbs: **a, b**, MB.Am.1207 right forelimb; **c, d**, MB.Am.1124 left forelimb, note the advanced differentiation of the radius compared to the ulna. *Sclerocephalus* forelimb: **e, f**, MB.Am.1314 right forelimb. *Micromelerpeton* hind limbs: **g, h**, MB.Am.1105 left hind limb; **i, j**, MB.Am.1341 right hind limb; **k, l**, MB.Am.1159 left hind limb; **m, n**, MB.Am.1207 right hind limb; **o, p**, MB.Am.1124 right hind limb. *Sclerocephalus* hind limbs: **q, r**, MB.Am.1305 hind limbs (left leg is on the left, right leg is on the right); **s, t**, MB.Am.1314 right hind limb; **u, v**, MB.Am.1313 right hind limb. Elements in black are fully ossified; stippling indicates the degree of ossification; asterisks indicate preaxial side of limb. **w**, Left hind leg of an extant salamander (*Cryptobranchus*, FMNH 814154) cleared and stained for bone (red) and cartilage (blue) showing ossification proceeding from pre- to postaxial. **x**, Bilateral asymmetry in the forelimbs of *Microbrachis*, cast of specimen MB.Am.822; the upper radius is crushed. Note the well-developed scapula and differentiated state of upper zeugopodial elements. Zeugopods in the bottom part of the figure lag behind in development, are undifferentiated and lack an olecranon process. F, femur; fi, fibula; H, humerus; mc, metacarpals; mt, metatarsals; ph, phalanges; ra, radius; sc, scapula; ti, tibia; ul, ulna. Scale bars, 1 mm.

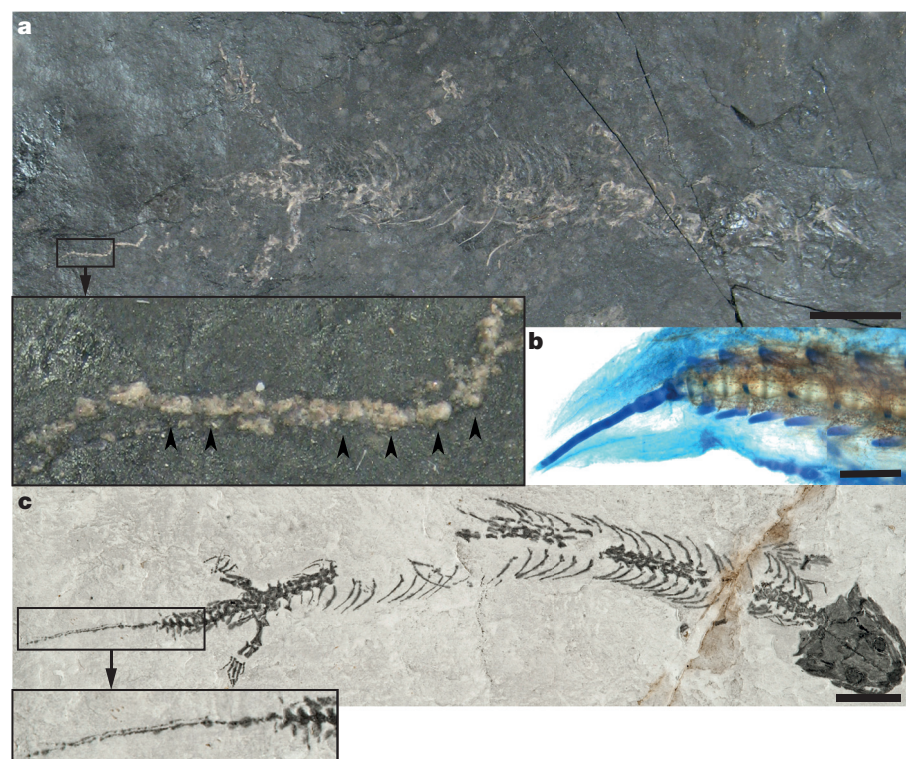
branchiosaurids were capable of regenerating their limbs, *Micro-melerpeton* were shown to possess this capacity on the basis of a unique pattern and combination of abnormalities associated with limb regeneration in modern salamanders<sup>6</sup>. Interestingly, new data about the molecular control of limb regeneration show that salamander orphan genes are involved in regeneration and probably also plays a role in the preaxial patterning during limb development, indicating a possible link between these two features in modern salamanders<sup>4,5</sup>. The data from the fossil record show that both preaxial polarity and the capacity to regenerate limbs were definitely present at the base of dissorophoid temnospondyls 290 million years ago, but may even have evolved earlier at the base of Tetrapoda (Fig. 1).

Furthermore, novel ontogenetic data from the lepospondyl lineage of Tetrapoda provide important data points for an even deeper picture of the evolution of limb development and regenerative capacities in tetrapods. The available ontogenetic data of two microsaurian taxa, *Microbrachis* and *Hyloplezion*, from the Upper Carboniferous coal deposits of Nýrany, Czech Republic, suggest that, contrary to the temnospondyl taxa, these ‘microsaurs’ probably displayed postaxial polarity in limb development as indicated by a delayed differentiation of the preaxial zeugopodial and metapodial elements<sup>21</sup>. Notably, a number of *Microbrachis* specimens show a clear asymmetry in ossification between the left and right limbs<sup>21</sup>. The degree of ossification and morphological differentiation of limb elements on one side is well progressed and in accordance with the overall developmental stage of the specimen. However, on the other side, the proximal limb elements (scapula and humerus) are equally well developed, while more distal limb elements show a much less mature state of differentiation (Fig. 2x). This suggests that the distal limbs of these microsaurs specimens may be in the process of regenerating their lower limbs, leading to the obvious disjunction of differentiation between proximal and distal elements.

Even stronger support for salamander-like regenerative capacities in lepospondyl microsaurs comes from clear evidence of tail regeneration in *Microbrachis* and *Hyloplezion*. While tail regeneration in these taxa has been previously mentioned in brief<sup>7,21</sup>, here we show that the regeneration of vertebral elements during tail regeneration in

microsaurs proceeded in a salamander-like fashion (Fig. 3a–c). In extant vertebrates, tail regeneration is known to occur in lungfish<sup>22</sup>, salamanders<sup>23,24</sup>, frog tadpoles<sup>25</sup> and lizards<sup>26</sup>. Tail regeneration in tadpoles is only possible until metamorphic climax and differs from salamander tail regeneration in many respects, including tissue organization and the molecular markers involved<sup>25</sup>. Moreover, tadpoles lack ossified vertebral structures in the tail, which is resorbed entirely during metamorphosis. Unfortunately, very little is known about lungfish tail and fin regeneration<sup>22</sup>. Among amniotes, many lizards are known to autotomize their tail as a defence mechanism when threatened by a predator<sup>26,27</sup>. While the subsequently regenerated tail is functional as a balancing organ, articulated vertebrae are replaced by an unsegmented hyaline cartilage tube<sup>26,27</sup>. In contrast, salamanders form a blastema at the tail amputation site, similar to the blastema initiating limb regeneration<sup>23,28</sup>, and fully regenerate vertebral elements as well as the associated musculature and innervation<sup>23,24</sup>. Initially, a long rod of cartilage forms, which is already subdivided on the cellular level into regions of hyaline cartilage cells with abundant interstitial matrix and interpolated regions of fusiform cartilage cells lacking matrix<sup>24</sup>. These segmented cartilages subsequently differentiate further and represent the precocious centra of the regenerated tail<sup>23</sup>. The sequence in which vertebral elements regenerate is reversed compared to original axial development in that centra form first, followed by the musculature and finally neural and haemal arches<sup>23,24</sup>.

In the microsaurs *Microbrachis* and *Hyloplezion*, the precocious centra of the regenerating tail vertebrae are clearly visible (Fig. 3a, c), showing further differentiation from an unsegmented rod into segmented precocious centra. *Microbrachis* specimen MB.Am.815.6 (Fig. 3c) shows a line of small, segmented elements, which attach to differentiated caudal vertebrae at the former amputation site. In *Hyloplezion* (specimen St.209, Fig. 3a), a clear differentiation between dense precocious centra and interspersed hyaline zones is visible in a pattern notably similar to regenerating tails of modern salamanders<sup>23</sup>. The degree of differentiation suggests a more advanced stage of tail regeneration in *Hyloplezion* specimen St.209 than in *Microbrachis* specimen MB.Am.815.6, although taphonomic and taxonomic variation may play a role as well. A further *Hyloplezion* specimen (RSM.1899.32.3.)



**Figure 3 | Tail regeneration in the microsaurs taxa *Microbrachis* and *Hyloplezion* in comparison with an extant salamander.** **a**, *Hyloplezion* specimen St 209 showing tail regeneration; inset represents enlarged region of the regenerating tail of St 209, showing segmentation and differentiation into primordial centra (black arrowheads). **b**, Enlarged view of regenerating tail of a cleared and stained salamander (*Ambystoma tigrinum*) specimen (FMNH 281665). **c**, *Microbrachis* specimen MB.Am.815 showing tail regeneration. Inset shows close-up of regenerating section of the tail. Scale bars, 1 mm (a and c) and 2 mm (b).



also shows tail regeneration, but owing to poorer preservation a more detailed assessment of the stage of development of the central primordia is not possible. As in salamanders, haemal and neural arches are still lacking at the available stages of tail regeneration in the two lepospondyl taxa. This form of full tail regeneration is otherwise exclusively known from modern salamanders and, together with the indication for possible limb regeneration in 'microsaurs', provides evidence for salamander-like regenerative capacities previously unrecognized for lepospondyls.

Among modern tetrapods, only salamanders still conserve the capacity to regenerate their limbs and many other organs fully, including fully functional tails, throughout their entire lifespan. New molecular evidence indicates that salamander-specific genes are involved in the capacity to regenerate limbs in modern salamanders and are implicated in preaxial polarity in limb development<sup>4,5</sup>. The novel evidence from the fossil record suggests that regenerative capacities are an ancient feature of Tetrapoda or even Sarcopterygii (tetrapods and lobe-finned fish), and were lost at least once in the evolutionary history of tetrapods along the lineage leading to amniotes (Fig. 1). Tail regeneration and the potential ability of lepospondyls to regenerate limbs show that salamander-like regeneration was also possible in fossil tetrapod taxa that probably had postaxial polarity in limb development.

The definitive co-occurrence of preaxial polarity in limb development and the capacity to regenerate limbs at the base of the dissorophoid clade within Temnospondyli demonstrates that both features were already present together in the lineage leading to modern salamanders, at least 80 million years before the first occurrence of stem salamanders in the fossil record. Moreover, preaxial polarity in limb development and salamander-like regenerative capacities may be ancient features of tetrapod vertebrates that are conserved only in salamanders (Fig. 1).

Received 6 May; accepted 13 August 2015.

Published online 26 October 2015.

- Shubin, N., Tabin, C. & Carroll, S. B. Fossils, genes and the evolution of animal limbs. *Nature* **388**, 639–648 (1997).
- Zeller, R., Lopez-Rios, J. & Zuniga, A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nature Rev. Genet.* **10**, 845–858 (2009).
- Fröbisch, N. B. & Shubin, N. H. Salamander limb development: integrating genes, morphology, and fossils. *Dev. Dyn.* **240**, 1087–1099 (2011).
- Brookes, J. P. & Gates, P. Mechanisms underlying vertebrate limb regeneration: lessons from the salamander. *Biochem. Soc. Trans.* **42**, 625–630 (2014).
- Brookes, J. in *Salamanders in Regeneration Research: Methods and Protocols* (eds Kumar, A. & Andrés, S.) Ch. 1 (Springer, 2015).
- Fröbisch, N. B., Bickelmann, C. & Witzmann, F. Early evolution of limb regeneration in tetrapods: evidence from a 300-million-year-old amphibian. *Proc. Biol. Sci.*, (2014).
- Carroll, R. L. & Gaskill, P. *The Order Microsauria*. 122 (American Philosophical Society, 1978).
- Shubin, N. H. & Alberch, P. A morphogenetic approach to the origin and basic organisation of the tetrapod limb. *Evol. Biol.* **20**, 319–387 (1986).
- Shubin, N. H. & Wake, D. B. in *Amphibian Biology* Vol. 5 (eds Heatwole, H. & Davies, M.) 1782–1808 (Surrey Beatty & Sons PTY limited, 2003).
- Gardiner, D. M. & Bryant, S. V. in *Fins into Limbs* (ed. Hall, B. K.) 163–182 (Univ. Chicago Press, 2007).
- Simon, A. & Tanaka, E. M. Limb regeneration. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 291–300 (2013).
- Anderson, J. S., Reisz, R. R., Fröbisch, N. B., Scott, D. & Sumida, S. S. A stem batrachian from the Early Permian of Texas and the origin of frogs and salamanders. *Nature* **453**, 515–518 (2008).
- Ruta, M. & Coates, M. I. Dates, nodes and character conflict: addressing the lissamphibian origin problem. *J. Syst. Palaeontology* **5**, 69–122 (2007).
- Marjanović, D. & Laurin, M. The origin(s) of extant amphibians: a review with emphasis on the "lepospondyl hypothesis". *Geodiversitas* **35**, 207–272 (2013).
- Schoch, R. R. Life cycles, plasticity and palaeoecology in temnospondyl amphibians. *Palaeontology* **57**, 517–529 (2014).
- Fröbisch, N. B., Carroll, R. L. & Schoch, R. R. Limb ossification in the Paleozoic branchiosaurid *Apaton* (Temnospondyli) and the early evolution of preaxial dominance in tetrapod limb development. *Evol. Dev.* **9**, 69–75 (2007).
- Fröbisch, N. B. Ossification patterns in the tetrapod limb – conservation and divergence from morphogenetic events. *Biol. Rev. Camb. Philos. Soc.* **83**, 571–600 (2008).
- Schmalhausen, J. J. Die Entwicklung des Extremitätenskelettes von *Salamandrella kayserringii*. *Anat. Anz.* **37**, 431–446 (1910).
- Schoch, R. R. Early larval ontogeny of the Permo-Carboniferous temnospondyl *Scelerocephalus*. *Paleontology* **46**, 1055–1072 (2003).
- Witzmann, F. & Pfretzschner, H.-U. Larval ontogeny of *Micromelerpeton credneri* (Temnospondyli, Dissorophoidea). *J. Vert. Paleontol.* **23**, 750–768 (2003).
- Olori, J. Skeletal morphogenesis of *Microbrachis* and *Hylaplesion* (Tetrapoda: Lepospondyli), and implications for the developmental patterns of extinct, early tetrapods. *PLoS ONE* **10**, e0128333 (2015).
- Conant, E. B. Regeneration in the African lungfish, *Protopterus*. I. Gross aspects. *J. Exp. Zool.* **174**, 15–31 (1970).
- Holtzer, H., Holtzer, S. & Avery, G. An experimental analysis of the development of the spinal column IV. Morphogenesis of tail vertebrae during regeneration. *J. Morphol.* **96**, 145–171 (1955).
- Vaglia, J. L., Babcock, S. K. & Harris, R. N. Tail development and regeneration throughout the life cycle of the four-toed salamander *Hemidactylum scutatum*. *J. Morphol.* **233**, 15–29 (1997).
- Beck, C. W., Izpisua Belmonte, J. C. & Christen, B. Beyond early development: *Xenopus* as an emerging model for the study of regenerative mechanisms. *Dev. Dyn.* **238**, 1226–1248 (2009).
- Fisher, R. E. et al. A histological comparison of the original and regenerated tail in the green anole, *Anolis carolinensis*. *Anat. Rec.* **295**, 1609–1619 (2012).
- Alibardi, L. *Morphological and Cellular Aspects of Tail and Limb Regeneration in Lizards*. (Springer, 2010).
- Holtzer, S. W. The inductive activity of the spinal cord in urodele tail regeneration. *J. Morphol.* **99**, 1–39 (1956).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank B. Ekrt, J. Müller, U. Göhlich, S. Walsh, C. Mehling, P. Barrett and R. Schoch for access to the collections under their care. H. J. Götz took the photograph in Fig. 3c; S. Lokatis took the photograph in Fig. 3b. This research was funded by an DFG Emmy Noether Grant (FR 2647/5-1) to N.B.F.; the Jackson School of Geosciences, the Banks Fellowship, Society of Vertebrate Paleontology Estes Memorial Grant and Paleontological Society Lane Student Award to J.C.O.; and the Feodor-Lynen Fellowship of the Alexander von Humboldt Foundation to F.W.

**Author Contributions** N.B.F., C.B. and F.W. designed the research; N.B.F., C.B., J.C.O. and F.W. performed the research; N.B.F. wrote the manuscript; C.B., J.C.O. and F.W. contributed to the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.B.F. ([nadia.froebisch@mf-n-berlin.de](mailto:nadia.froebisch@mf-n-berlin.de)).



# Global non-linear effect of temperature on economic production

Marshall Burke<sup>1,2\*</sup>, Solomon M. Hsiang<sup>3,4\*</sup> & Edward Miguel<sup>4,5</sup>

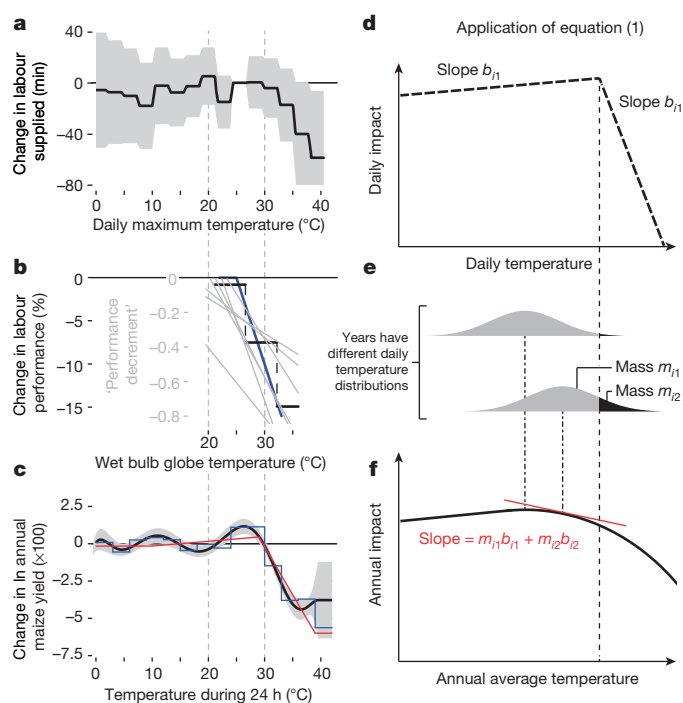
Growing evidence demonstrates that climatic conditions can have a profound impact on the functioning of modern human societies<sup>1,2</sup>, but effects on economic activity appear inconsistent. Fundamental productive elements of modern economies, such as workers and crops, exhibit highly non-linear responses to local temperature even in wealthy countries<sup>3,4</sup>. In contrast, aggregate macroeconomic productivity of entire wealthy countries is reported not to respond to temperature<sup>5</sup>, while poor countries respond only linearly<sup>5,6</sup>. Resolving this conflict between micro and macro observations is critical to understanding the role of wealth in coupled human–natural systems<sup>7,8</sup> and to anticipating the global impact of climate change<sup>9,10</sup>. Here we unify these seemingly contradictory results by accounting for non-linearity at the macro scale. We show that overall economic productivity is non-linear in temperature for all countries, with productivity peaking at an annual average temperature of 13 °C and declining strongly at higher temperatures. The relationship is globally generalizable, unchanged since 1960, and apparent for agricultural and non-agricultural activity in both rich and poor countries. These results provide the first evidence that economic activity in all regions is coupled to the global climate and establish a new empirical foundation for modelling economic loss in response to climate change<sup>11,12</sup>, with important implications. If future adaptation mimics past adaptation, unmitigated warming is expected to reshape the global economy by reducing average global incomes roughly 23% by 2100 and widening global income inequality, relative to scenarios without climate change. In contrast to prior estimates, expected global losses are approximately linear in global mean temperature, with median losses many times larger than leading models indicate.

Economic productivity—the efficiency with which societies transform labour, capital, energy, and other natural resources into new goods or services—is a key outcome in any society because it has a direct impact on individual wellbeing. While it is well known that temperature affects the dynamics of virtually all chemical, biological and ecological processes, how temperature effects recombine and aggregate within complex human societies to affect overall economic productivity remains poorly understood. Characterizing this influence remains a fundamental problem both in the emerging field of coupled human–natural systems and in economics more broadly, as it has implications for our understanding of historical patterns of human development and for how the future economy might respond to a changing climate.

Prior analyses have identified how specific components of economic production, such as crop yields, respond to temperature using high-frequency micro-level data<sup>3,4</sup>. Meanwhile, macro-level analyses have documented strong correlations between total economic output and temperature over time<sup>5,6</sup> and across space<sup>13,14</sup>, but it is unknown whether these results are connected, and if so, how. In particular, strong responses of output to temperature observed in micro data from wealthy countries are not apparent in existing macro studies<sup>5</sup>. If

wealthy populations actually are unaffected by temperature, this could indicate that wealth and human-made capital are substitutes for natural capital (for example, the composition of the atmosphere) in economic activity<sup>5,7</sup>. Resolving this apparent discrepancy thus has central implications for understanding the nature of sustainable development<sup>7</sup>.

Numerous basic productive components of an economy display a highly non-linear relationship with daily or hourly temperature<sup>1</sup>. For example, labour supply<sup>4</sup>, labour productivity<sup>6</sup>, and crop yields<sup>3</sup> all decline abruptly beyond temperature thresholds located between 20 °C and 30 °C (Fig. 1a–c). However, it is unclear how these abrupt



**Figure 1 | Highly non-linear micro responses generate smooth and shifted macro response.** **a–c**, Highly non-linear micro-level responses of labour supply<sup>4</sup> (**a**), labour performance<sup>6</sup> (**b**) and crop yield<sup>3</sup> (**c**) to daily temperature exposure exhibit similar ‘kinked’ structures between 20 and 30 °C. **d**, **e**, These micro-level responses ( $f_i(T)$  in equation (1); **d**) map onto country-level distributions of temperatures across different locations and times within that country ( $g_i(T - \bar{T})$  in equation (1); **e**). Shifts in country-level distributions correspond to changes in average annual temperature, altering the fraction of unit-hours ( $m_{i1}$  and  $m_{i2}$ ) exposed to different regions of the micro-level response in **d**, **f**. Aggregating daily impacts according to equation (1) maps annual average temperature to annual output as a non-linear and concave function that is smoother than the micro response with a lower optimum ( $Y(\bar{T})$  in equation (1)).

<sup>1</sup>Department of Earth System Science, Stanford University, California 94305, USA. <sup>2</sup>Center on Food Security and the Environment, Stanford University, California 94305, USA. <sup>3</sup>Goldman School of Public Policy, University of California, Berkeley, California 94720, USA. <sup>4</sup>National Bureau of Economic Research. <sup>5</sup>Department of Economics, University of California, Berkeley, California, 94720, USA.

\*These authors contributed equally to this work.

declines at the micro level are reflected in coarser macro-level data. When production is integrated over large regions (for example, countries) or long units of time (for example, years), there is a broad distribution of momentary temperatures to which individual components of the economy (for example, crops or workers) are exposed. If only the hottest locations or moments cause abrupt declines in output, then when combined with many cooler and highly productive moments they would sum to an aggregate level of output that only declines modestly when aggregate average temperature increases.

To fix ideas, let function  $f_i(T)$  describe the productive contribution of an individual productive unit in industry  $i$  (for example, a firm) relative to instantaneous (for example, daily) temperature  $T$  (Fig. 1d). For a given country, period, and industry, denote the fraction of unit-hours spent below the critical temperature threshold as  $m_{i1}$  and the fraction above as  $m_{i2}$  (Fig. 1e). The full distribution of unit-hours across all temperatures is  $g_i(T - \bar{T})$ , centred at average temperature  $\bar{T}$ . Assume  $g_i(\cdot)$  is mean zero. If productivity loss within a single productive unit-hour has limited impact on other units, as suggested by earlier findings<sup>8,15</sup>, then aggregate production  $Y$  is the sum of output across industries, each integrated over all productive unit-hours in the country and period:

$$Y(\bar{T}) = \sum_i Y_i(\bar{T}) = \sum_i \int_{-\infty}^{\infty} f_i(T) \cdot g_i(T - \bar{T}) dT \quad (1)$$

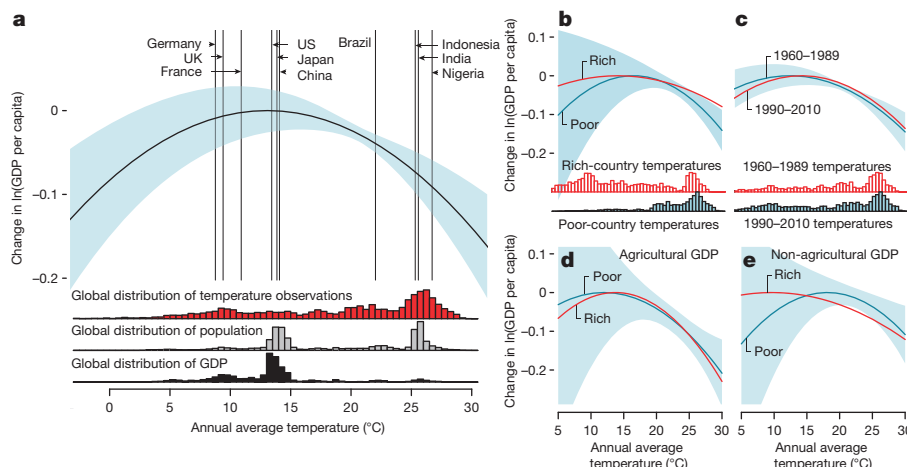
As  $\bar{T}$  rises and a country warms on average,  $m_{i2}$  increases gradually for all productive units (Fig. 1e). This growing number of hours beyond the temperature threshold imposes gradual but increasing losses on total output  $Y(\bar{T})$ .

Equation (1) predicts that  $Y(\bar{T})$  is a smooth concave function (Fig. 1f) with a derivative that is the average derivative of  $f_i(T)$  weighted by the number of unit-hours in each industry at each daily temperature. It also predicts that  $Y(\bar{T})$  peaks at a temperature lower than the threshold value in  $f_i(T)$ , if the slope of  $f_i(T)$  above the threshold is steeper than minus the slope below the threshold, as suggested by micro-scale evidence. These predictions differ fundamentally from notions that macro responses should closely mirror highly non-linear micro responses<sup>6,16</sup>. Importantly, while aggregate productivity losses ought to occur contemporaneous with temperature changes, these changes might also influence the long-run trajectory of an economy's output<sup>5,15</sup>. This could occur, for example, if temporary contemporaneous losses alter the rate of investment in new productive units, thereby altering future production. See Supplementary Equations 1–14 for details.

We test these predictions using data on economic production<sup>17</sup> for 166 countries over the period 1960–2010. In an ideal experiment, we would compare two identical countries, warm the temperature of one and compare its economic output to the other. In practice, we can approximate this experiment by comparing a country to itself in years when it is exposed to warmer- versus cooler-than-average temperatures<sup>18</sup> due to naturally occurring stochastic atmospheric changes. Heuristically, an economy observed during a cool year is the ‘control’ for that same society observed during a warmer ‘treatment’ year. We do not compare output across different countries because such comparisons are probably confounded, distinguishing our approach from cross-sectional studies that attribute differences across countries to their temperatures<sup>13</sup>.

We estimate how economic production changes relative to the previous year—that is, annual economic growth—to purge the data of secular factors in each economy that evolve gradually<sup>5</sup>. We deconvolve economic growth to account for: (1) all constant differences between countries, for example, culture or history; (2) all common contemporaneous shocks, for example, global price changes or technological innovations; (3) country-specific quadratic trends in growth rates, which may arise, for example, from changing political institutions or economic policies; and (4) the possibly non-linear effects of annual average temperature and rainfall. This approach is more reliable than only adjusting for observed variables because it accounts for unobserved time-invariant and time-trending covariates, allows these covariates to influence different countries in different ways, and outperforms alternative models along numerous dimensions<sup>15</sup> (see Supplementary Information). In essence, we analyse whether country-specific deviations from growth trends are non-linearly related to country-specific deviations from temperature and precipitation trends, after accounting for any shocks common to all countries.

We find country-level economic production is smooth, non-linear, and concave in temperature (Fig. 2a), with a maximum at 13 °C, well below the threshold values recovered in micro-level analyses and consistent with predictions from equation (1). Cold-country productivity increases as annual temperature increases, until the optimum. Productivity declines gradually with further warming, and this decline accelerates at higher temperatures (Extended Data Fig. 1a–g). This



**Figure 2 | Effect of annual average temperature on economic production.** **a**, Global non-linear relationship between annual average temperature and change in log gross domestic product (GDP) per capita (thick black line, relative to optimum) during 1960–2010 with 90% confidence interval (blue, clustered by country,  $N = 6,584$ ). Model includes country fixed effects, flexible trends, and precipitation controls (see Supplementary Methods). Vertical lines indicate average temperature for selected countries, although averages

are not used in estimation. Histograms show global distribution of temperature exposure (red), population (grey), and income (black). **b**, Comparing rich (above median, red) and poor (below median, blue) countries. Blue shaded region is 90% confidence interval for poor countries. Histograms show distribution of country-year observations. **c**, Same as **b** but for early (1960–1989) and late (1990–2010) subsamples (all countries). **d**, Same as **b** but for agricultural income. **e**, Same as **b** but for non-agricultural income.

result is globally representative and not driven by outliers (Extended Data Fig. 1h). It is robust to estimation procedures that allow the response of countries to change as they become richer (Extended Data Fig. 1i and Supplementary Table 1), use higher-order polynomials or restricted cubic splines to model temperature effects (Extended Data Fig. 1j–k), exclude countries with few observations, exclude major oil producers, exclude China and the United States, account for continent-specific annual economic shocks<sup>19</sup>, weaken assumptions about trends in growth, account for multiple lags of growth, and use alternative economic data sources<sup>20</sup> (Extended Data Table 1).

Accounting for delayed effects of temperature, which might be important if countries ‘catch up’ after temporary losses, increases statistical uncertainty but does not alter the net negative average effect of hot temperatures (Extended Data Fig. 2a–c). This ‘no catch up’ behaviour is consistent with the observed response to other climatological disturbances, such as tropical cyclones<sup>15</sup>.

While much of global economic production is clustered near the estimated temperature optimum (Fig. 2a, black histogram), both rich and poor countries exhibit similar non-linear responses to temperature (Fig. 2b). Poor tropical countries exhibit larger responses mainly because they are hotter on average, not because they are poorer (Extended Data Fig. 1i and Supplementary Table 1). There is suggestive evidence that rich countries might be somewhat less affected by temperature, as previously hypothesized<sup>5</sup>, but their response is statistically indistinguishable from poor countries at all temperatures (Extended Data Fig. 2d–f and Extended Data Table 2). Although the estimated total effect of high temperatures on rich countries is substantially less certain because there are few hot, rich countries in the sample, the non-linearity of the rich-country response alone is statistically significant ( $P < 0.1$ ; Extended Data Table 2), and we estimate an 80% likelihood that the marginal effect of warming is negative at high temperatures in these countries (Extended Data Fig. 2m). Our finding that rich countries respond non-linearly to temperature is consistent with recent county-level results in the United States<sup>8</sup>.

Our non-linear results are also consistent with the prior finding of no linear correlation between temperature and growth in rich countries<sup>5</sup>. Because the distribution of rich-country temperatures is roughly symmetrical about the optimum, linear regression recovers no association. Accounting for non-linearity reconciles this earlier result (Extended Data Fig. 3a and Supplementary Table 3) but reverses how wealth and technology are understood to mediate economic responses to temperature.

We do not find that technological advances or the accumulation of wealth and experience since 1960 has fundamentally altered the relationship between productivity and temperature. Results using data from 1960–1989 and 1990–2010 are nearly identical (Fig. 2c). In agreement with recent micro-level evidence<sup>8,21</sup>, substantial observed warming over the period apparently did not induce notable adaptation.

Consistent with micro-level findings that both agricultural and non-agricultural labour-related productivity are highly non-linear in instantaneous temperature<sup>3,4,6</sup>, we find agricultural and non-agricultural aggregate production are non-linear in average annual temperature for both rich and poor countries (Fig. 2d, e and Extended Data Fig. 2g–i). Low temperature has no significant effect on these subsamples, although limited poor-country exposure to these temperatures severely limits statistical precision. High temperatures have significant negative effects in all cases for poor countries, and significant or marginally significant effects for rich countries (Extended Data Fig. 2p–u).

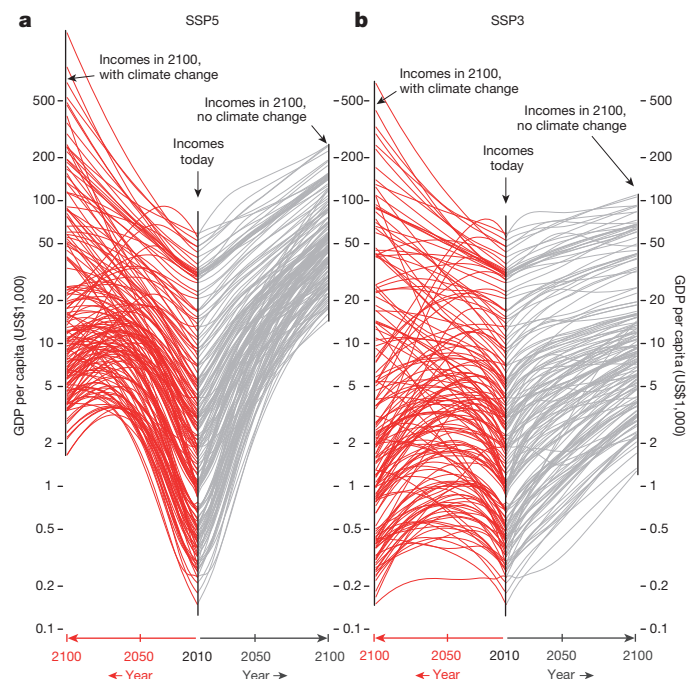
A global non-linear response of economic production to annual temperature has important implications for the likely economic impact of climate change. We find only weak suggestive evidence that richer populations are less vulnerable to warming, and no evidence that experience with high temperatures or technological advances since 1960 have altered the global response to temperature. This suggests that adaptation to climatic change may be more difficult than

previously believed<sup>9,10</sup>, and that the accumulation of wealth, technology and experience might not substantially mitigate global economic losses during this century<sup>8,21</sup>.

We quantify the potential impact of warming on national and global incomes by combining our estimated non-linear response function with ‘business as usual’ scenarios (Representative Concentration Pathway (RCP)8.5) of future warming and different assumptions regarding future baseline economic and population growth<sup>22</sup> (see Supplementary Information). This approach assumes future economies respond to temperature changes similarly to today’s economies—perhaps a reasonable assumption given the observed lack of adaptation during our 50-year sample.

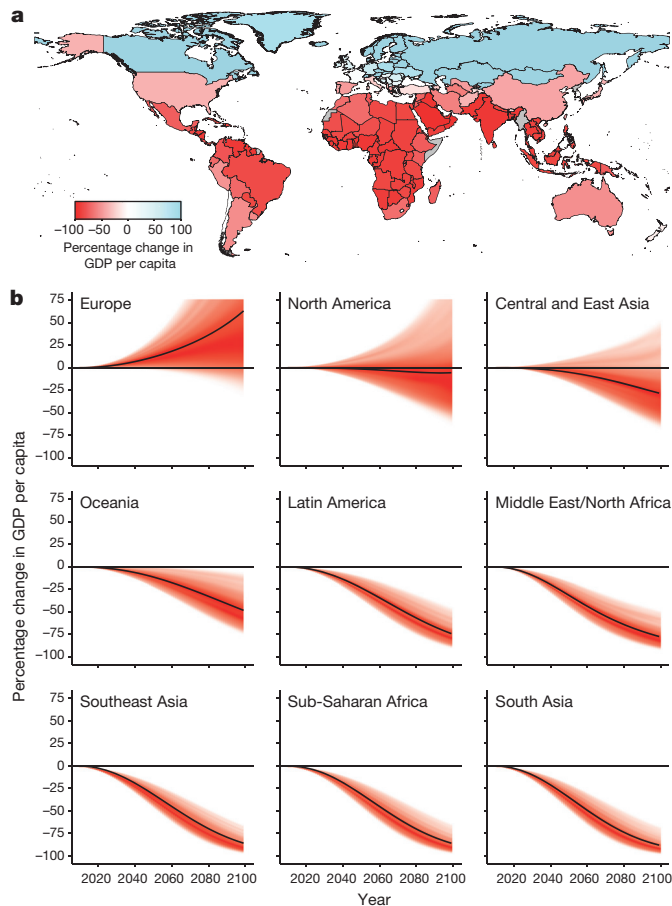
In 2100, we estimate that unmitigated climate change will make 77% of countries poorer in per capita terms than they would be without climate change. Climate change may make some countries poorer in the future than they are today, depending on what secular growth rates are assumed. With high baseline growth and unmitigated climate change (RCP8.5 and Shared Socio-economic Pathway (SSP)5; see Supplementary Information), we project that 5% of countries are poorer in 2100 than today (Fig. 3a), while with low growth, 43% are (SSP3; Fig. 3b).

Differences in the projected impact of warming are mainly a function of countries’ baseline temperatures, since warming raises productivity in cool countries (Fig. 4). In particular, Europe could benefit from increased average temperatures. Because warming harms productivity in countries with high average temperatures, incomes in poor regions are projected to fall relative to a world without climate change with high confidence ( $P < 0.01$ ), regardless of the statistical approach used. Models allowing for delayed effects project more negative impacts in colder wealthy regions; projections assuming rich and poor countries respond differently (Fig. 2b) are more uncertain because fewer data are used to estimate each response (Extended Data Fig. 4).



**Figure 3 | Country-level income projections with and without temperature effects of climate change.** **a, b,** Projections to 2100 for two socioeconomic scenarios<sup>22</sup> consistent with RCP8.5 ‘business as usual’ climate change: **a**, SSP5 assumes high baseline growth and fast income convergence; **b**, SSP3 assumes low baseline growth and slow convergence. Centre in each panel is 2010, each line is a projection of national income. Right (grey) are incomes under baseline SSP assumptions, left (red) are incomes accounting for non-linear effects of projected warming.

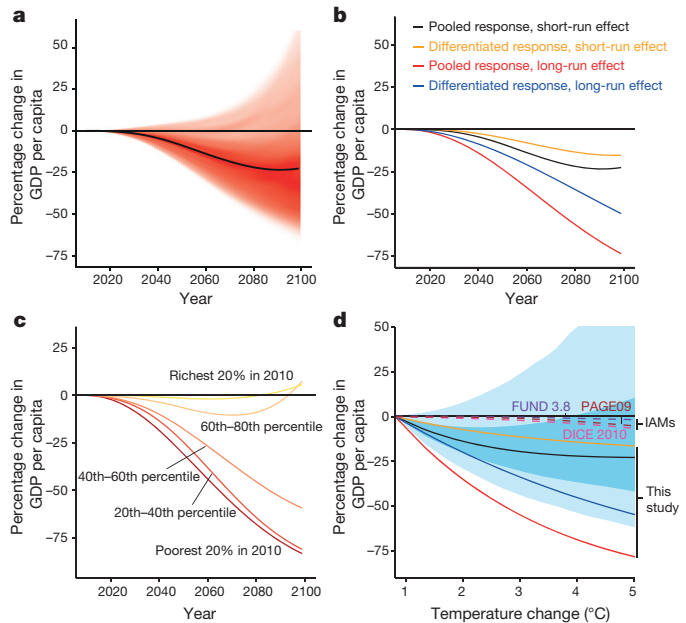




**Figure 4 | Projected effect of temperature changes on regional economies.** **a**, **b**, Change in GDP per capita (RCP8.5, SSP5) relative to projection using constant 1980–2010 average temperatures. **a**, Country-level estimates in 2100. **b**, Effects over time for nine regions. Black lines are projections using point estimates. Red shaded area is 95% confidence interval, colour saturation indicates estimated likelihood an income trajectory passes through a value<sup>27</sup>. Base maps by ESRI.

The impact of warming on global economic production is a population-weighted average of country-level impacts in Fig. 4a. Using our benchmark model (Fig. 2a), climate change reduces projected global output by 23% in 2100 (best estimate, SSP5) relative to a world without climate change, although statistical uncertainty allows for positive impacts with probability 0.29 (Fig. 5a and Extended Data Table 3). Estimates vary in magnitude, but not in structure, depending on the statistical approach (Fig. 5b and Extended Data Table 3). Models with delayed impacts project larger losses because cold countries gain less, while differentiated rich–poor models have smaller losses (statistical uncertainty allows positive outcomes with probability 0.09–0.40). Models allowing both delayed impacts and differentiated rich–poor responses (the most flexible approach) project global losses 2.2 times larger than our benchmark approach. In all cases, the likelihood of large global losses is substantial: global losses exceed 20% of income with probability 0.44–0.87 (Extended Data Table 3 and Extended Data Fig. 5).

Accounting for the global non-linear effect of temperature is crucial to constructing income projections under climate change because countries are expected to become both warmer and richer in the future. In a previous analysis in which a linear relationship was assumed and no significant linear effect was observed in rich countries<sup>5</sup>, it was hypothesized that countries adapted effectively to temperature as they became wealthier. Under this hypothesis, the impacts of future warming should lessen over time as countries become richer. In contrast,



**Figure 5 | Global damage estimates arising from non-linear effects of temperature.** **a**, Change in global GDP by 2100 using benchmark model (Fig. 2a). Calculation and display are the same as Fig. 4. **b**, Same as **a** (point estimate only) comparing approaches to estimating temperature effects (pooled/differentiated: rich and poor countries assumed to respond identically/differently, respectively; short run/long run: effects account for 1 or 5 years of temperature, respectively; see Supplementary Methods). **c**, Mean impacts by 2010 income quintile (benchmark model). **d**, Projected income loss in 2100 (SSP5) for different levels of global mean temperature increase, relative to pre-industrial temperatures. Solid lines marked as in **b**. Blue shaded areas are interquartile range and 5th–95th percentile estimates. Dashed lines show corresponding damages from major integrated assessment models (IAMs)<sup>12</sup>.

when we account for the non-linear effect of temperature historically, we find that rich and poor countries behave similarly at similar temperatures, offering little evidence of adaptation. This indicates that we cannot assume rich countries will be unaffected by future warming, nor can we assume that the impacts of future warming will attenuate over time as countries become wealthier. Rather, the impact of additional warming worsens over time as countries become warmer. As a result, projections using linear and non-linear approaches diverge substantially—by roughly 50–200% in 2100 (Extended Data Fig. 3c, d)—highlighting the importance of accounting for this non-linearity when assessing the impacts of future warming.

Strong negative correlation between baseline income and baseline temperature indicates that warming may amplify global inequality because hot, poor countries will probably suffer the largest reduction in growth (Fig. 5c). In our benchmark estimate, average income in the poorest 40% of countries declines 75% by 2100 relative to a world without climate change, while the richest 20% experience slight gains, since they are generally cooler. Models with delayed impacts do not project as dramatic differences because colder countries also suffer large losses (Extended Data Fig. 5).

We use our results to construct an empirical ‘damage function’ that maps global temperature change to global economic loss by aggregating country-level projections. Damage functions are widely used in economic models of global warming, but previously relied on theory for structure and rough estimates for calibration<sup>11,12</sup>. Using our empirical results, we project changes to global output in 2100 for different temperature changes (Fig. 5d; see Supplementary Information) and compare these to previously estimated damage functions<sup>12</sup>. Commonly used functions are within our estimated uncertainty, but differ in two important respects.

First, our projected global losses are roughly linear—and slightly concave—in temperature, not quadratic or exponential as previously theorized. Approximate linearity results from the broad distribution of temperature exposure within and across countries, which causes the country-weighted average derivative of the productivity function in Fig. 2a to change little as countries warm and prevents abrupt transitions in global output even though the contribution of individual productive units are highly non-linear (see Fig. 1). Global losses are slightly concave in global temperature because the effect of compounding negative growth declines mechanically over time (Extended Data Fig. 6e and Supplementary Information). These properties are independent of the growth scenario and response function (Extended Data Fig. 6a).

Second, the slope of the damage function is large even for slight warming, generating expected costs of climate change 2.5–100 times larger than prior estimates for 2 °C warming, and at least 2.5 times larger for higher temperatures (Extended Data Fig. 6b–d). Notably, our estimates are based only on temperature effects (or effects for which historical temperature has been a proxy), and so do not include other potential sources of economic loss associated with climate change, such as tropical cyclones<sup>15</sup> or sea-level rise<sup>23</sup>, included in previous damage estimates.

If societies continue to function as they have in the recent past, climate change is expected to reshape the global economy by substantially reducing global economic output and possibly amplifying existing global economic inequalities, relative to a world without climate change. Adaptations such as unprecedented innovation<sup>24</sup> or defensive investments<sup>25</sup> might reduce these effects, but social conflict<sup>2</sup> or disrupted trade<sup>26</sup>—either from political restrictions or correlated losses around the world—could exacerbate them.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 April; accepted 15 September 2015.**

**Published online 21 October 2015.**

1. Dell, M., Jones, B. F. & Olken, B. A. What do we learn from the weather? The new climate-economy literature. *J. Econ. Lit.* **52**, 740–798 (2014).
2. Hsiang, S. M., Burke, M. & Miguel, E. Quantifying the influence of climate on human conflict. *Science* **341**, 1235367 (2013).
3. Schlenker, W. & Roberts, M. J. Non-linear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proc. Natl Acad. Sci. USA* **106**, 15594–15598 (2009).
4. Graff Zivin, J. & Neidell, M. Temperature and the allocation of time: Implications for climate change. *J. Labor Econ.* **13**, 1–26 (2014).
5. Dell, M., Jones, B. F. & Olken, B. A. Climate change and economic growth: evidence from the last half century. *Am. Econ. J. Macroecon.* **4**, 66–95 (2012).
6. Hsiang, S. M. Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proc. Natl Acad. Sci. USA* **107**, 15367–15372 (2010).

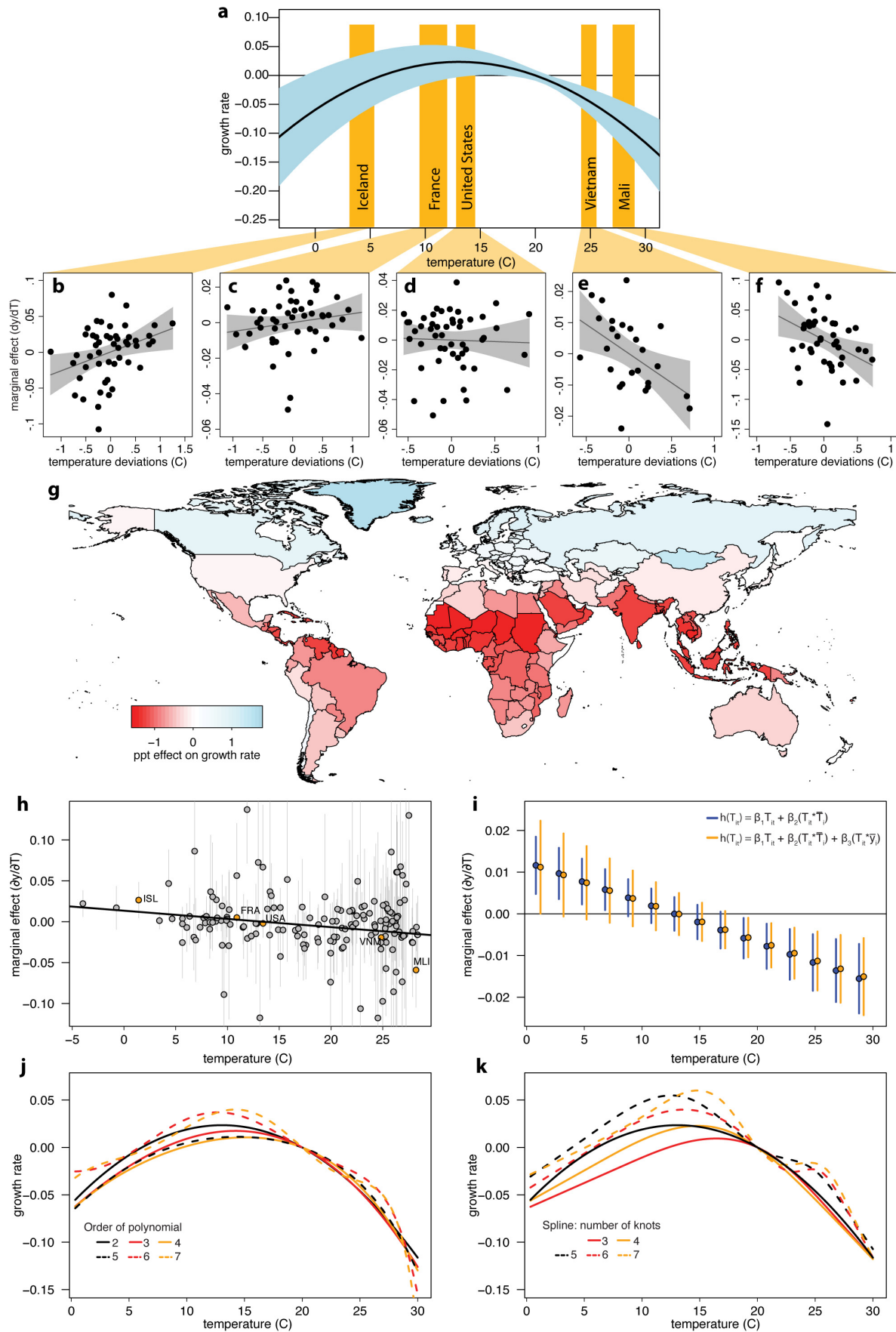
7. Solow, R. in *Economics of the Environment* (ed. Stavins, R.) (W. W. Norton & Company, 2012).
8. Deryugina, T. & Hsiang, S. M. Does the environment still matter? Daily temperature and income in the United States. *NBER Working Paper* 20750 (2014).
9. Tol, R. S. J. The economic effects of climate change. *J. Econ. Perspect.* **23**, 29–51 (2009).
10. Nordhaus, W. A *Question of Balance: Weighing the Options on Global Warming Policies* (Yale Univ. Press, 2008).
11. Pindyck, R. S. Climate change policy: what do the models tell us? *J. Econ. Lit.* **51**, 860–872 (2013).
12. Revesz, R. L. et al. Global warming: improve economic models of climate change. *Nature* **508**, 173–175 (2014).
13. Nordhaus, W. D. Geography and macroeconomics: new data and new findings. *Proc. Natl Acad. Sci. USA* **103**, 3510–3517 (2006).
14. Dell, M., Jones, B. F. & Olken, B. A. Temperature and income: reconciling new cross-sectional and panel estimates. *Am. Econ. Rev.* **99**, 198–204 (2009).
15. Hsiang, S. M. & Jina, A. The causal effect of environmental catastrophe on long run economic growth. *NBER Working Paper* 20352 (2014).
16. Heal, G. & Park, J. Feeling the heat: temperature, physiology & the wealth of nations. *NBER Working Paper* 19725 (2013).
17. World Bank Group. *World Development Indicators 2012* (World Bank Publications, 2012).
18. Matsuura, K. & Willmott, C. J. Terrestrial air temperature and precipitation: monthly and annual time series (1900–2010) v. 3.01. [http://climate.geog.udel.edu/~climate/html\\_pages/README.ghcn\\_ts2.html](http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html) (2012).
19. Hsiang, S. M., Meng, K. C. & Cane, M. A. Civil conflicts are associated with the global climate. *Nature* **476**, 438–441 (2011).
20. Summers, R. & Heston, A. The Penn World Table (Mark 5): an expanded set of international comparisons, 1950–1988. *Q. J. Econ.* **106**, 327–368 (1991).
21. Burke, M. & Emerick, K. Adaptation to climate change: evidence from US agriculture. *Am. Econ. J. Econ. Pol.* (in the press).
22. O'Neill, B. C. et al. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Clim. Change* **122**, 387–400 (2014).
23. Houser, T. et al. *Economic Risks of Climate Change: An American Prospectus* (Columbia Univ. Press, 2015).
24. Olmstead, A. L. & Rhode, P. W. Adapting North American wheat production to climatic challenges, 1839–2009. *Proc. Natl Acad. Sci. USA* **108**, 480–485 (2011).
25. Barreca, A., Clay, K., Deschenes, O., Greenstone, M. & Shapiro, J. S. Adapting to climate change: the remarkable decline in the US temperature-mortality relationship over the 20th century. *J. Polit. Econ.* (in the press).
26. Costinot, A., Donaldson, D. & Smith, C. Evolving comparative advantage and the impact of climate change in agricultural markets: evidence from a 9 million-field partition of the earth. *J. Polit. Econ.* (in the press).
27. Hsiang, S. M. Visually-weighted regression. *SSRN Working Paper* 2265501 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank D. Anthoff, M. Auffhammer, V. Bosetti, M. P. Burke, T. Carleton, M. Dell, L. Goulder, S. Heft-Neal, B. Jones, R. Kopp, D. Lobell, F. Moore, J. Rising, M. Tavoni, and seminar participants at Berkeley, Harvard, Princeton, Stanford universities, Institute for the Study of Labor, and the World Bank for useful comments.

**Author Contributions** M.B. and S.M.H. conceived of and designed the study; M.B. and S.M.H. collected and analysed the data; M.B., S.M.H. and E.M. wrote the paper.

**Author Information** Replication data have been deposited at the Stanford Digital Repository (<http://purl.stanford.edu/wb587wt4560>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. ([mburke@stanford.edu](mailto:mburke@stanford.edu)).



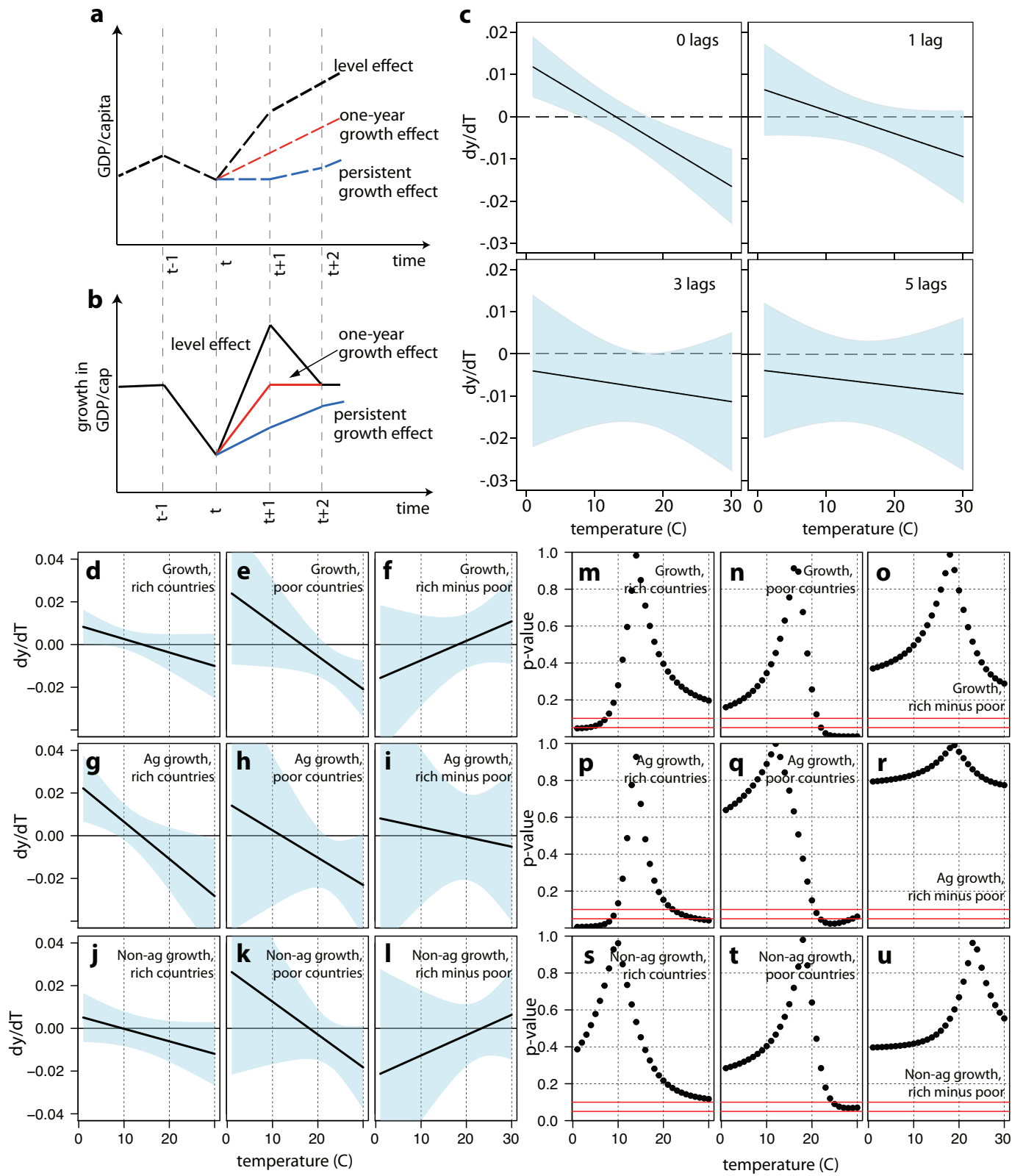


**Extended Data Figure 1 | Understanding the non-linear response function.**

**a**, Response function from Fig. 2a. **b–f**, The global non-linear response reflects changing marginal effects of temperature at different mean temperatures. Plots represent selected country-specific relationships between temperature and growth over the sample period, after accounting for the controls in Supplementary Equation (15); dots are annual observations for each country, dark line the estimated linear relationship, grey area the 95% confidence interval. **g**, Percentage point effect of uniform 1°C warming on country-level growth rates, as estimated using the global relationship shown in **a**. A value of  $-1$  indicates that a country growing at  $3\% \text{ yr}^{-1}$  during the baseline period is projected to grow at  $2\% \text{ yr}^{-1}$  with  $+1^\circ\text{C}$  warming. ppt, percentage point. **h**, Dots represent estimated marginal effects for each country from separate linear time-series regressions (analogous to slopes of lines in **b–f**), and grey lines the 95% confidence interval on each. The dark black line plots the derivative  $\frac{\partial Y}{\partial T}$  of the estimated global response function in Fig. 2a. **i**, Global

non-linearity is driven by differences in average temperature, not income. Blue dots (point estimates) and lines (95% confidence interval) show marginal effects of temperature on growth evaluated at different average temperatures, as estimated from a model that interacts country annual temperature with country average temperature (see Supplementary Equation (17);  $\frac{\partial Y}{\partial T_{it}} = \hat{\beta}_1 + \hat{\beta}_2 \cdot \bar{T}_i$ ).

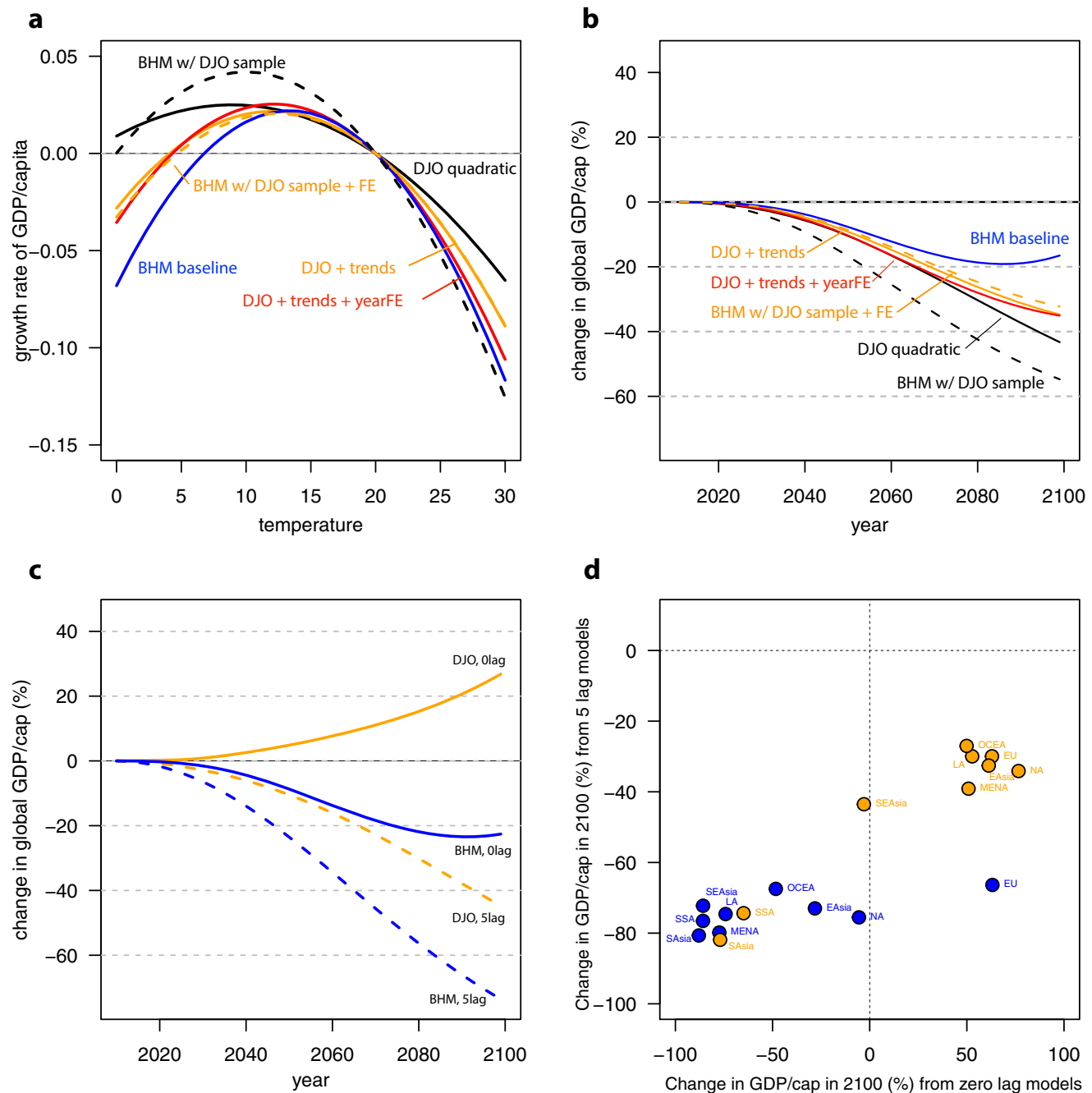
Orange dots and lines show equivalent estimates from a model that includes an interaction between annual temperature and average GDP. Point estimates are similar across the two models, indicating that the non-linear response is not simply due to hot countries being poorer on average. **j–k**, More flexible functional forms yield similar non-linear global response functions. **j**, Higher-order polynomials in temperature, up to order 7. **k**, Restricted cubic splines with up to 7 knots. Solid black line in both plots is quadratic polynomial shown in **a**. Base maps by ESRI.



**Extended Data Figure 2 | Growth versus level effects, and comparison of rich and poor responses.** **a**, Evolution of GDP per capita given a temperature shock in year  $t$ . Black line shows a level effect, with GDP per capita returning to its original trajectory immediately after the shock. Red line shows a 1-year growth effect, and blue line a multi-year growth effect. **b**, Corresponding pattern in the growth in per-capita GDP. Level effects imply a slower-than-average growth rate in year  $t$  but higher-than-average rate in  $t + 1$ . Growth effects imply lower rates in year  $t$  and then average rates thereafter (for a 1-year shock) or lower rates thereafter (if a 1-year shock has persistent effects on growth). **c**, Cumulative marginal effect of temperature on growth as additional lags are included; solid line indicates the sum of the contemporaneous and lagged marginal effects at a given temperature level, and the blue areas its 95% confidence interval. **d–l**, Testing the null that slopes of rich- and poor-country response functions are zero, or the same as one another, for quadratic

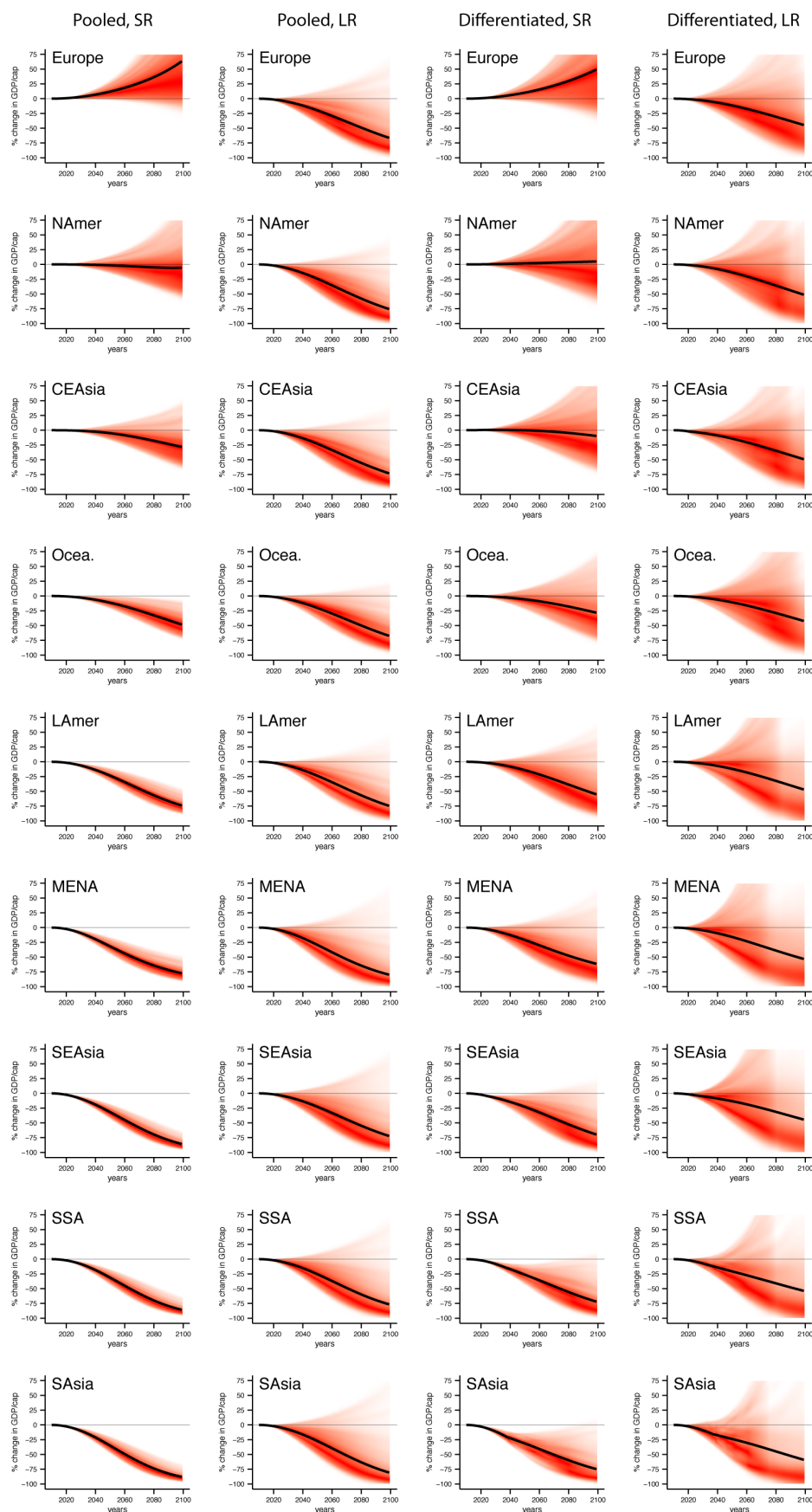
response functions shown in Fig. 2. Black lines show the point estimate for the marginal effect of temperature on rich-country production for different initial temperatures (blue shading is 95% confidence interval) (**d**, **g**, **j**), the marginal effect poor-country production for different initial temperatures (**e**, **h**, **k**), and the estimated difference between the marginal effect on rich- and poor-country production compared at each initial temperature (**f**, **i**, **l**). **d–f**, Effects on economy-wide per-capita growth (corresponding to Fig. 2b). **g–i**, Agricultural growth. **j–l**, Non-agricultural growth. **m–u**, Corresponding  $P$  values. Each point represents the  $P$  value on the test of the null hypothesis that the slope of the rich-country response is zero at a given temperature (**m**, **p**, **s**), that the slope of the poor-country response is zero (**n**, **q**, **t**), or that rich- and poor-country responses are equal (**o**, **r**, **u**) for overall growth, agricultural growth, or non-agricultural growth, respectively. **m–u**, Red lines at the bottom of each plot indicate  $P = 0.10$  and  $P = 0.05$ .





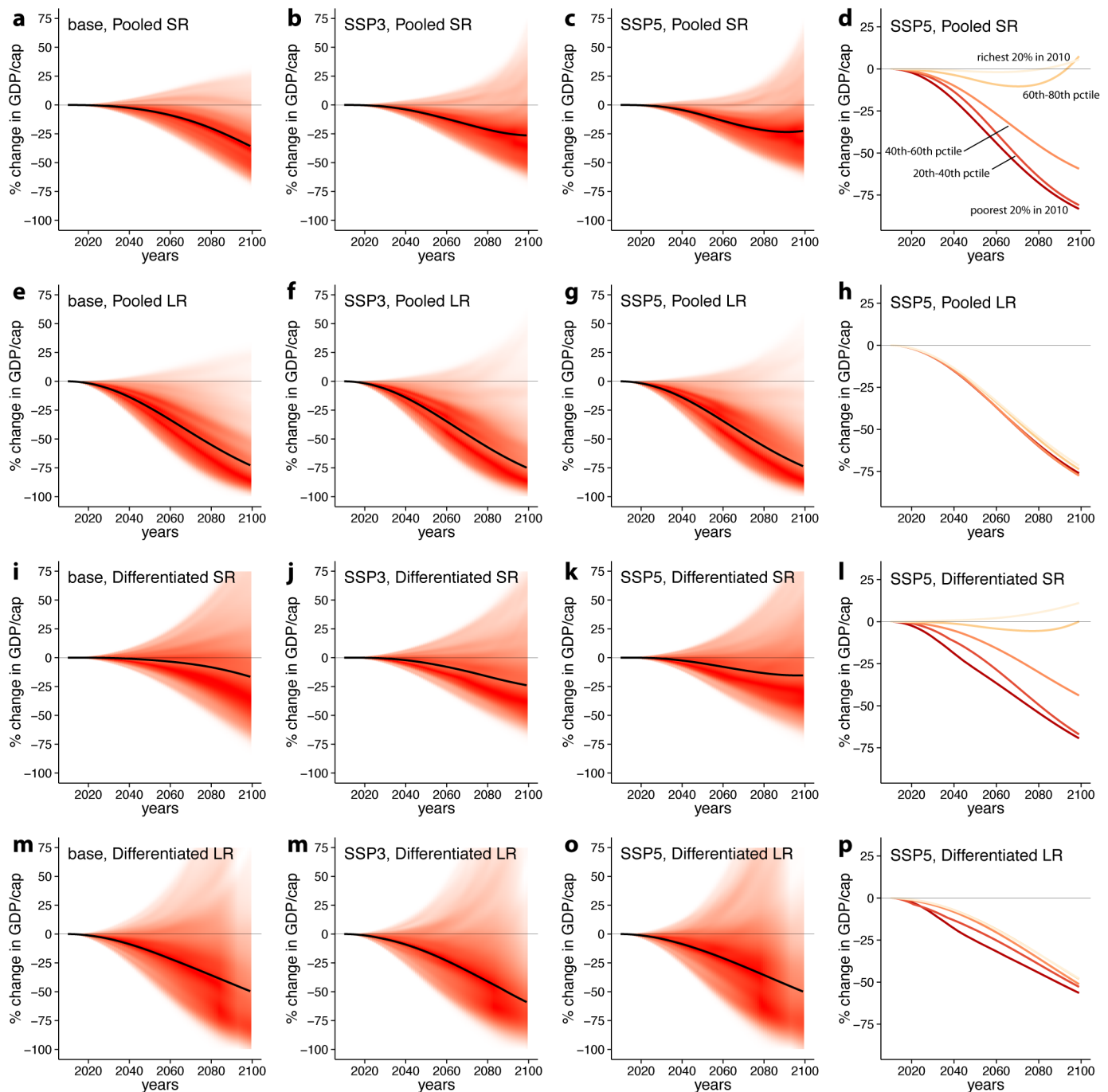
**Extended Data Figure 3 | Comparison of our results and those of Dell, Jones and Olken<sup>5</sup>.** **a**, Allowing for non-linearity in the original Dell, Jones and Olken (DJO)<sup>5</sup> data/analysis indicates a similar temperature–growth relationship as in our results (BHM) under various choices about data sample and model specification (coefficients in Supplementary Table 3). **b**, Projections of future global impacts on per-capita GDP (RCP8.5, SSP5) using the re-estimated non-linear DJO response functions in **a** again provide similar estimates to our baseline BHM projection (shown in blue, and here using the sample of countries with > 20 years of data to match the DJO preferred sample).

**c**, Projected global impacts differ substantially between DJO and BHM if DJO's original linear results are used to project impacts. Lines show projected change in global GDP per capita by 0- and 5-lag pooled non-linear models in BHM (blue), and 0- and 5-lag linear models in DJO (orange). **d**, Projected regional impacts also differ strongly between BHM's non-linear and DJO's linear approach. Plot shows projected impacts on GDP per capita in 2100 by region, for the 0-lag model (x-axis) and 5-lag model (y-axis), with BHM estimates in blue and DJO estimates in orange. See Supplementary Discussion for additional detail.



**Extended Data Figure 4 | Projected impact of climate change (RCP8.5, SSP5) on regional per capita GDP by 2100, relative to a world without climate change, under the four alternative historical response functions. Pooled short-run (SR) response (column 1), pooled long-run (LR) response**

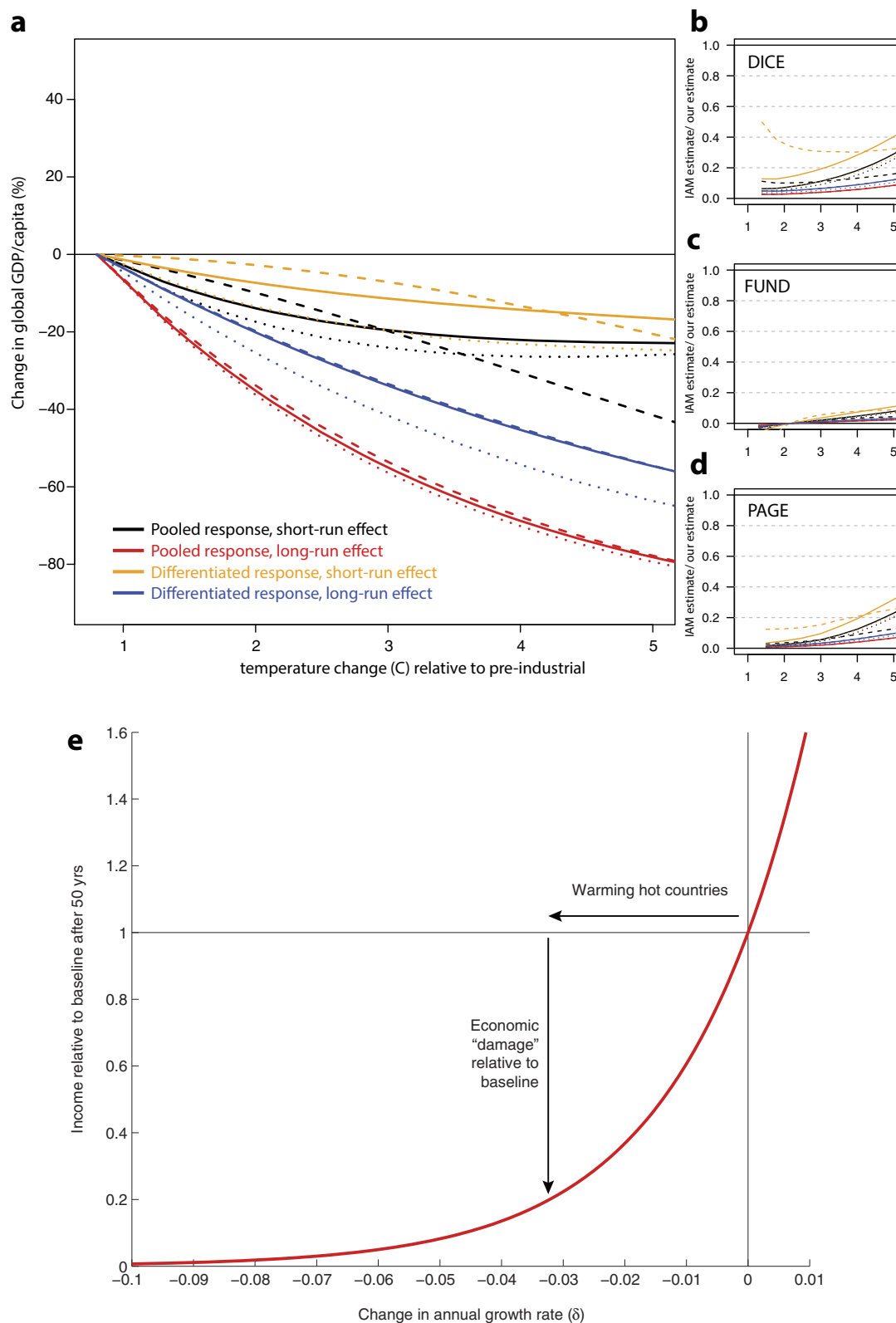
(column 2), differentiated SR response (column 3), differentiated LR response (column 4). Shading is as in Fig. 5a. CEAsia, Central and East Asia; LAmer, Latin America; MENA, Middle East/North Africa; NAmér, North America; Ocea, Oceania; SAsia, South Asia; SEAsia, South-east Asia; SSA, sub-Saharan Africa.



**Extended Data Figure 5 | Projected impact of climate change (RCP8.5) by 2100 relative to a world without climate change, for different historical response functions and different future socioeconomic scenarios. a–p.** The first three columns show impacts on global per-capita GDP (analogous to Fig. 5a), for the three different underlying socioeconomic scenarios and four different response functions shown in Fig. 5b. Last column (d, h, l, p) shows

impact on per capita GDP by baseline income quintile (as in Fig. 5c), for SSP5 and the different response functions. Colours correspond to the income quintiles as labelled in d. Globally aggregated impact projections are more sensitive to choice of response function than projected socioeconomic scenario, with response functions that allow for accumulating effects of temperature (LR) showing more negative global impacts but less inequality in these impacts.





**Extended Data Figure 6 | Estimated damages at different levels of temperature increase by socioeconomic scenario and assumed response function, and comparison of these results to damage functions in IAMs.**

**a**, Percentage loss of global GDP in 2100 under different levels of global temperature increase, relative to a world in which temperatures remained at pre-industrial levels (as in Fig. 5d). Colours indicated in figure represent different historical response functions (as in Fig. 5b). Line type indicates the underlying assumed socioeconomic scenario: dash indicates ‘base’ (United Nations medium variant population projections, future growth rates are

country-average rates observed 1980–2010), dots indicate SSP3, solid lines indicate SSP5. **b–d**, The ratio of estimated damages from each IAM using data from ref. 12 (shown in Fig. 5d) to damages in **a**. Colours as in **a** for results from this study; IAM results are fixed across scenarios and response functions. Temperature increase is in °C by 2100, relative to pre-industrial levels. **e**, Explanation for why economic damage function is concave: increasingly negative growth effects have diminishing cumulative impact in absolute levels over finite periods (see Supplementary Discussion). Red curve is  $e^{\delta\zeta}$  after  $\zeta = 50$  years.

**Extended Data Table 1 | Regression estimates for global sample, main estimate and robustness**

	(1) Base	(2) >20yrs	(3) No oil	(4) No US/China	(5) ContYr FE	(6) ContYr + noTrend	(7) No YrFE	(8) LinearTime	(9) LDV 1lag	(10) LDV 3lags	(11) PWT
Temp.	0.0127*** (0.0038)	0.0135*** (0.0038)	0.0128*** (0.0036)	0.0128*** (0.0038)	0.0142*** (0.0037)	0.0133*** (0.0034)	0.0103*** (0.0039)	0.0128*** (0.0043)	0.0087** (0.0039)	0.0062* (0.0038)	0.0072* (0.0039)
Temp. sq.	-0.0005*** (0.0001)	-0.0005*** (0.0001)	-0.0005*** (0.0001)	-0.0005*** (0.0001)	-0.0005*** (0.0001)	-0.0004*** (0.0001)	-0.0004*** (0.0001)	-0.0005*** (0.0001)	-0.0004*** (0.0001)	-0.0003*** (0.0001)	-0.0004*** (0.0001)
Precip.	0.0145 (0.0100)	0.0148 (0.0100)	0.0130 (0.0101)	0.0148 (0.0101)	0.0124 (0.0106)	0.0084 (0.0098)	0.0159 (0.0107)	0.0137 (0.0100)	0.0165* (0.0095)	0.0201** (0.0098)	0.0195* (0.0109)
Precip. sq.	-0.0047* (0.0026)	-0.0049* (0.0026)	-0.0040 (0.0025)	-0.0048* (0.0026)	-0.0041 (0.0027)	-0.0021 (0.0023)	-0.0045* (0.0026)	-0.0035 (0.0023)	-0.0047* (0.0024)	-0.0052** (0.0024)	-0.0038 (0.0028)
Constant	1.4575** (0.6444)	0.0740 (0.0633)	1.4522** (0.6228)	1.4707** (0.6507)	-0.0362 (0.0411)	-0.0819** (0.0370)	-0.8024** (0.3366)	-0.7693*** (0.0517)	-11.3227*** (0.7957)	-28.3451*** (2.0763)	0.0643 (0.0467)
Observations	6584	6477	6090	6484	6584	6584	6584	6584	6418	6086	6627
R squared	0.286	0.278	0.275	0.284	0.367	0.267	0.240	0.219	0.286	0.289	0.220
Optimum	13.06	13.39	12.64	13.09	14.92	17.40	12.74	13.40	11.92	9.98	9.88

Unless otherwise indicated, all models include country fixed effects, year fixed effects, and quadratic country time trends, with errors clustered at the country level. Temperature is measured in °C and precipitation in metres. Columns: (1) main specification, (2) as in column 1 but excluding countries with fewer than 20 years of growth data, (3) as in column 1 but dropping large oil exporting countries, (4) as in column 1 but dropping United States and China, (5) as in column 1 but adding continent-by-year fixed effects, (6) as in column 1 but adding continent-by-year fixed effects and dropping country time trends, (7) as in column 1 but dropping year fixed effects, (8) as in column 1 but only linear time trend, (9–10) as in column 1 but adding 1 or 3 lags of per capita growth (that is, lagging the dependent variable), (11) as in column 1 but using growth data from Penn World Tables. Asterisks indicate statistical significance at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels.

**Extended Data Table 2 | Comparing temperature effects on per-capita growth in rich versus poor countries**

	(1) Base	(2) poor-yr FE	(3) >20Yrs	(4) >20Yrs+poor-yr FE	(5) ContYr FE	(6) ContYr + noTrend
Temperature ( $\beta_1$ )	0.0089** (0.0044)	0.0069 (0.0050)	0.0085* (0.0044)	0.0063 (0.0050)	0.0099* (0.0054)	0.0146*** (0.0052)
Temperature sq. ( $\beta_2$ )	-0.0003* (0.0002)	-0.0003 (0.0002)	-0.0003 (0.0002)	-0.0003 (0.0002)	-0.0003 (0.0002)	-0.0005*** (0.0002)
Temperature * 1[poor] ( $\beta_3$ )	0.0165 (0.0182)	0.0193 (0.0187)	0.0246 (0.0186)	0.0275 (0.0191)	0.0115 (0.0147)	0.0016 (0.0147)
Temperature sq. * 1[poor] ( $\beta_4$ )	-0.0005 (0.0004)	-0.0005 (0.0004)	-0.0006 (0.0004)	-0.0007 (0.0004)	-0.0004 (0.0004)	0.0001 (0.0003)
Observations	6452	6452	6345	6345	6452	6452
R squared	0.291	0.299	0.284	0.293	0.372	0.270
$\beta_1 + \beta_3$	0.0254	0.0262	0.0332	0.0338	0.0215	0.0162
se( $\beta_1 + \beta_3$ )	0.0177	0.0180	0.0181	0.0184	0.0130	0.0123
$\beta_2 + \beta_4$	-0.0008	-0.0008	-0.0009	-0.0009	-0.0007	-0.0004
se( $\beta_2 + \beta_4$ )	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003
Rich-country optimum (C)	14.1	12.2	14.2	12.2	19.2	15.1
Poor-country optimum (C)	16.5	17.2	17.8	18.3	16.2	20.6

Unless otherwise indicated, all models include a quadratic in precipitation, country fixed effects, year fixed effects, and quadratic country time trends, with errors clustered at the country level. Temperature is measured in °C. Columns: (1) main specification with the climate variables interacted with an indicator for whether a country is poor, (2) as in column 1 but allowing the year fixed effects to differ across rich and poor countries, (3) as in column 1 but restricting sample to countries with at least 20 observations, (4) as in column 1 but restricting sample to countries with at least 20 observations and allowing year fixed effects to differ across rich and poor countries, (5) as in column 1 but adding continent-by-year fixed effects, (6) as in column 1 but adding continent-by-year fixed effects and dropping country time trends (as in ref. 5). The estimated linear and quadratic effects in poor countries (and their standard errors) are given in the bottom rows of the table, along with the estimated temperature optima for rich and for poor countries. Asterisks indicate statistical significance at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels for coefficients in the main part of the table.



**Extended Data Table 3 | Projected impacts of climate change on global GDP per capita by 2100 under RCP8.5, relative to a world without climate change**

Historical model	Statistic	Projected impacts under alternate scenarios of future population/income growth <i>% impact relative to world without climate change</i>		
		base	SSP3	SSP5
Pooled model	point est.	-36	-26	-23
	5th	-65	-60	-58
	25th	-50	-42	-39
	50th	-35	-24	-21
	75th	-16	2	5
	95th	27	67	66
	% runs < 0	85	73	71
	% runs < -10	79	66	63
	% runs < -20	71	56	51
Pooled with 5 lags	point est.	-73	-75	-74
	5th	-94	-94	-94
	25th	-85	-85	-84
	50th	-70	-68	-67
	75th	-48	-41	-41
	95th	28	55	51
	% runs < 0	91	89	89
	% runs < -10	89	86	87
	% runs < -20	87	84	84
Rich/poor model	point est.	-16	-24	-15
	5th	-67	-62	-61
	25th	-43	-42	-37
	50th	-17	-19	-12
	75th	22	16	27
	95th	103	82	101
	% runs < 0	63	66	60
	% runs < -10	55	59	51
	% runs < -20	48	50	44
Rich/poor with 5 lags	point est.	-50	-59	-50
	5th	-93	-90	-92
	25th	-75	-69	-72
	50th	-40	-39	-30
	75th	20	45	65
	95th	265	379	428
	% runs < 0	69	66	62
	% runs < -10	65	63	59
	% runs < -20	61	59	56

Estimates are from four different response functions (as in Fig. 5b) estimating how growth responds to temperature, and three different scenarios of how future populations and incomes will evolve without climate change. The left column indicates the historical response function on which the projections are based, the second column describes the statistic being reported (either the point estimate, a given percentile in the bootstrapped distribution of projections, or the percentage of total runs projecting impacts more negative than zero, -10%, or -20%), and the last three columns give percentage impacts for three different future scenarios: 'base' (United Nations medium variant population projections, future growth rates without climate change equal to country-average rates observed 1980–2010), SSP3, and SSP5.

# Oxygen regulation of breathing through an olfactory receptor activated by lactate

Andy J. Chang<sup>1</sup>, Fabian E. Ortega<sup>1</sup>, Johannes Riegler<sup>2†</sup>, Daniel V. Madison<sup>3</sup> & Mark A. Krasnow<sup>1</sup>

**Animals have evolved homeostatic responses to changes in oxygen availability that act on different timescales. Although the hypoxia-inducible factor (HIF) transcriptional pathway that controls long-term responses to low oxygen (hypoxia) has been established<sup>1</sup>, the pathway that mediates acute responses to hypoxia in mammals is not well understood. Here we show that the olfactory receptor gene *Olfir78* is highly and selectively expressed in oxygen-sensitive glomus cells of the carotid body, a chemosensory organ at the carotid artery bifurcation that monitors blood oxygen and stimulates breathing within seconds when oxygen declines<sup>2</sup>. *Olfir78* mutants fail to increase ventilation in hypoxia but respond normally to hypercapnia. Glomus cells are present in normal numbers and appear structurally intact, but hypoxia-induced carotid body activity is diminished. Lactate, a metabolite that rapidly accumulates in hypoxia and induces hyperventilation<sup>3–6</sup>, activates *Olfir78* in heterologous expression experiments, induces calcium transients in glomus cells, and stimulates carotid sinus nerve activity through *Olfir78*. We propose that, in addition to its role in olfaction, *Olfir78* acts as a hypoxia sensor in the breathing circuit by sensing lactate produced when oxygen levels decline.**

The carotid body is the major sensor of blood oxygen in mammals. It is stimulated by a reduction in arterial blood oxygen from 100 mmHg to <80 mmHg (1 mmHg = 133.3 Pa). A current model is that hypoxia causes closure of K<sup>+</sup> channels of glomus cells, stimulating Ca<sup>2+</sup>-dependent release of neurotransmitters onto afferent nerves that signal to brainstem respiratory centres. However, the oxygen sensor and sensing mechanism that trigger these events in glomus cells remain controversial<sup>2</sup> (Extended Data Fig. 1a–d). To identify new candidate molecules involved in carotid body oxygen sensing, we used RNA sequencing and whole-genome microarrays to compare gene expression of the carotid body from wild-type adult C57BL/6J mice with that of the adrenal medulla, which shares developmental and functional similarities with the carotid body but does not respond acutely to hypoxia<sup>7</sup>. We reasoned that an oxygen sensor would be expressed at high levels in carotid body relative to adrenal medulla, and focused on signalling molecules that can act on the acute timescale of carotid body sensing. Transcripts for the olfactory receptor *Olfir78* were highly expressed in carotid body (top 4% of all genes by RNA sequencing) and highly enriched relative to adrenal medulla by both RNA sequencing (92-fold) and microarrays (three probe sets, 17- to 80-fold) (Fig. 1a, b, Extended Data Fig. 2a–d, and Extended Data Table 1).

Olfactory receptors (ORs) comprise a subfamily of G-protein-coupled receptors that is the largest gene family in vertebrates, encoded by ~1,200 genes in mouse<sup>8</sup>. ORs are expressed in olfactory sensory neurons and detect volatile odorants in smell. However, some ORs are expressed in other tissues<sup>8–10</sup>. The RNA sequencing results showed that three other OR genes (*Olfir1033*, *Olfir613*, and *Olfir856ps1*) were expressed (reads per kilobase per million (RPKM) > 2) in the carotid body, but at similar levels in adrenal medulla and thus not pursued

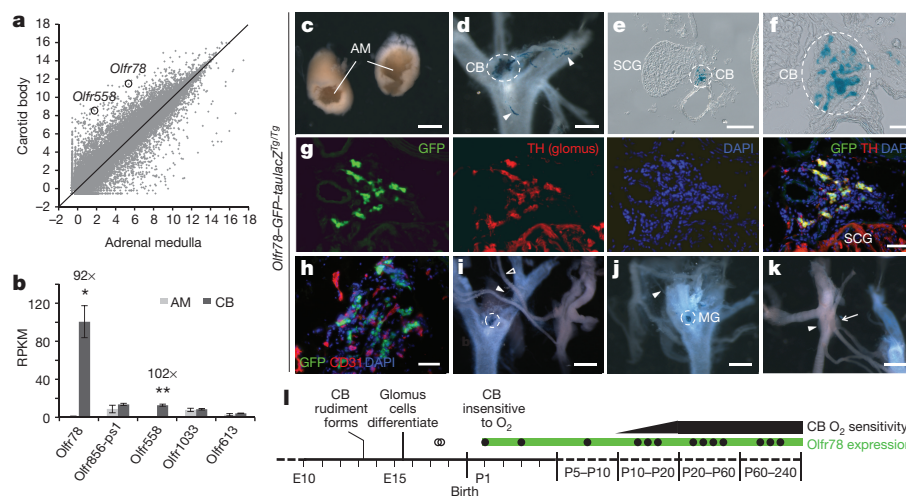
further (Fig. 1b and Extended Data Fig. 2b). *Olfir558* was highly and selectively expressed in the carotid body, but at lower levels than *Olfir78* (Fig. 1a, b). *Olfir78* and *Olfir558* encode closely related proteins of the same OR subfamily and lie in close proximity in the genome (Extended Data Fig. 2e). Because of the high and selective expression of *Olfir78* and *Olfir558* in the carotid body, we investigated their expression and potential function.

The carotid body is composed of type I glomus cells that sense changes in oxygen, type II sustentacular cells that resemble neuroglia, nerve fibres, and endothelial and smooth muscle cells that comprise fine tortuous vessels off the carotid artery (Extended Data Fig. 1b). To determine which cells express *Olfir78*, we used an *Olfir78* reporter strain carrying green fluorescent protein (GFP) and *taulacZ* genes in the 3'-untranslated region (3'-UTR) of the *Olfir78* locus<sup>11</sup>. X-gal staining for lacZ activity in adults confirmed strong and selective *Olfir78* expression in carotid body and no detectable adrenal gland expression (Fig. 1c–f). The cluster-like pattern of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) staining in carotid body suggested *Olfir78* is expressed in glomus cells (Fig. 1d–f). This was verified by antibody staining showing co-localization of *Olfir78* reporter GFP with tyrosine hydroxylase (TH), a glomus cell marker (Fig. 1g, h); 98% of all GFP- and TH-positive cells expressed both markers (*n* = 3 sections from 3 animals, 222 cells). Unlike the monoallelic expression of ORs observed in olfactory neurons<sup>12</sup>, we found that in animals carrying only one allele of the *Olfir78* reporter gene, 98% of all GFP- and TH-positive cells still expressed both markers (*n* = 3 sections from 3 animals, 271 cells, *P* = 0.461 Tg/+ versus Tg/Tg by unpaired *t*-test) (Extended Data Fig. 3a). Using an *Olfir558 lacZ* knock-in reporter allele, we detected reporter activity in some vascular smooth muscle cells of carotid body blood vessels, but not in glomus cells (Extended Data Fig. 3e–h). Thus we focused on *Olfir78*.

Although *Olfir78* and its human orthologue *OR51E2* are expressed in other tissues outside the olfactory system<sup>9,10,13,14</sup>, no expression of *Olfir78* reporter was detected in other parts of the oxygen-sensing circuit for breathing besides carotid body (Extended Data Fig. 1c): carotid sinus and glossopharyngeal nerves (Fig. 1i–k), petrosal, nodose/jugular, and superior cervical ganglia (Fig. 1i–k and Extended Data Fig. 3b, c), and brainstem (Extended Data Fig. 3d). In some carotid bifurcations, there were patches of *Olfir78*-expressing cells on arteries that were innervated by branches of the glossopharyngeal nerve distinct from the carotid sinus nerve (Fig. 1j); these may be ectopic 'miniglomera' with chemosensory functions similar to carotid body<sup>15</sup>. We conclude that *Olfir78* is specifically expressed in acute oxygen-sensing cells of the carotid body and not in afferent pathways or the respiratory centres themselves.

Because ORs mediate acute sensory signalling in olfaction, we tested whether *Olfir78* was involved in acute oxygen sensing in the carotid body by examining breathing of *Olfir78* knockout mice<sup>16</sup>. Homozygous *Olfir78*<sup>−/−</sup> mutants were viable and present in Mendelian ratios at

<sup>1</sup>Department of Biochemistry, Stanford University School of Medicine and Howard Hughes Medical Institute, Stanford, California 94305-5307, USA. <sup>2</sup>Department of Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>3</sup>Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>†</sup>Present address: Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA.

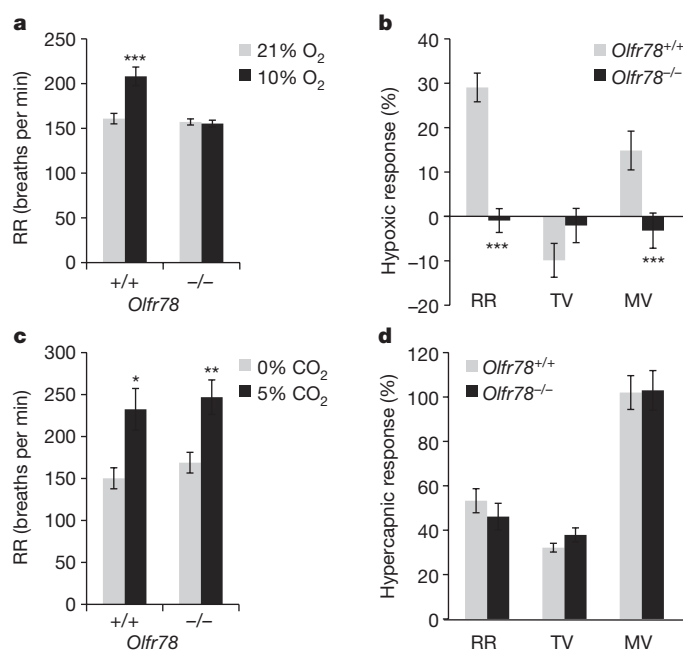


**Figure 1 | *Olfir78* is expressed in carotid body glomus cells.** **a**, Expression of 26,728 genes in adult mouse carotid body (CB) and adrenal medulla (AM) by RNA sequencing. Axes are log<sub>2</sub> values of number of aligned reads per 10<sup>7</sup> aligned reads generated. **b**, OR genes highly expressed in carotid body and/or adrenal medulla;  $\times$ , fold enrichment (carotid body/adrenal medulla). **a, b**, Sample size  $n = 3$  cohorts of 10 animals each. Data as mean (**a**) or mean  $\pm$  s.e.m. (**b**). \* $P < 0.05$ , \*\* $P < 0.01$  by paired  $t$ -test by cohort. **c-l**, Expression of *Olfir78* knock-in reporter mouse<sup>11</sup>. **c-f**, X-gal staining (blue) detects taulacZ ( $\beta$ -galactosidase) reporter activity. **c**, Adrenal gland showing adrenal medulla. Reporter not expressed. **d**, Carotid bifurcation (dorsal view, superior cervical ganglion (SCG) removed). Reporter expressed in carotid body (dashed circle) and sporadic blood vessels (arrowhead). **e, f**, Transverse section of carotid bifurcation (**e**) and

close-up (**f**). **g, h**, Immunostaining of carotid body sections. *Olfir78* reporter expression (GFP; green) co-localized with carotid body glomus cell marker (TH; red; **g**) but not endothelial cell marker (CD31; red; **h**). TH is also expressed in nerve fibres and SCG. DAPI (4',6-diamidino-2-phenylindole; blue), nuclei. **i-k**, X-gal stained carotid bifurcations (ventral view). **i**, Carotid body (dashed circle) innervated by carotid sinus nerve (filled arrowhead), a branch of glossopharyngeal nerve (open arrowhead). **j**, 'Miniglomerulus' (MG; dashed circle) innervated by glossopharyngeal nerve (arrowhead). **k**, Petrosal ganglion (arrow), nodose/jugular ganglia (arrowhead). Reporter not expressed. **l**, *Olfir78* reporter expression (X-gal staining) during carotid body development. Filled circles, robust expression; open circles, not detected. E, embryonic day. Scale bars, 500  $\mu$ m (**c-e, i-k**) and 100  $\mu$ m (**f-h**).

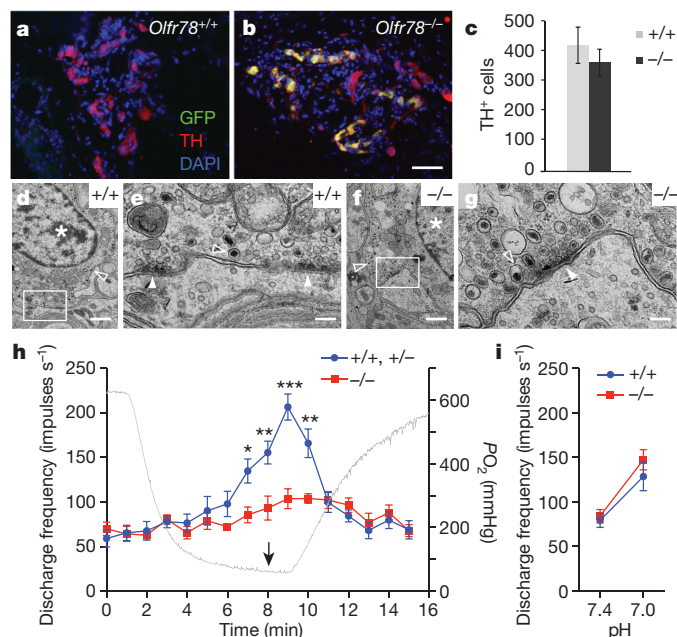
birth (postnatal day 1 (P1); 15:35:19 for  $+/+;+/-;-/-$  respectively;  $\chi^2 = 0.4783$ ,  $P > 0.7$ ) and weaning (P21; 50:115:64 for  $+/+;+/-;-/-$  respectively;  $\chi^2 = 1.7162$ ,  $P > 0.3$ ), and they appeared to breathe and behave normally under ambient conditions. However, when challenged by hypoxia (10% O<sub>2</sub>), *Olfir78*<sup>+/+</sup> control animals increased respiratory rate and minute ventilation, whereas *Olfir78*<sup>-/-</sup> mutants did not exhibit significant ventilatory changes (Fig. 2a, b and Extended Data Fig. 4a, b). Most strikingly, the respiratory rate of *Olfir78*<sup>-/-</sup> mutants did not change in hypoxia, while increasing ~30% in controls (Fig. 2a, b). In hypoxia, arterial blood from *Olfir78*<sup>-/-</sup> mutants had higher partial pressure of carbon dioxide (PaCO<sub>2</sub>) and lower pH than wild-type animals (Extended Data Fig. 5a-f), consistent with their inability to increase ventilation. By contrast, ventilatory responses to hypercapnia (5% CO<sub>2</sub>) remained intact in *Olfir78*<sup>-/-</sup> mutants (Fig. 2c, d and Extended Data Fig. 4c, d), as did two other rapid behavioural responses to hypoxia: reduced locomotion and more regular breathing (Supplementary Video 1). We also did not detect differences between controls and *Olfir78*<sup>-/-</sup> mutants in body temperature or metabolism in response to hypoxia (Extended Data Fig. 5g-j), parameters that can affect hypoxic ventilation in small mammals<sup>17</sup>. Thus, *Olfir78*<sup>-/-</sup> mutants have a specific defect in hypoxic regulation of respiratory rate, a physiological function controlled by the carotid body<sup>2</sup>.

Previous studies showed that mice with fewer carotid body glomus cells have attenuated responses to hypoxia<sup>18</sup>. We examined developmental expression of *Olfir78* in carotid body and found it was not expressed embryonically, when transcription factors that regulate carotid body development are detected and glomus cells form<sup>18</sup>. *Olfir78* expression was first observed after birth before maturation of carotid body oxygen sensing, and persisted throughout adult life (Fig. 1l). The number of TH-positive glomus cells and their organization into clusters were not affected in *Olfir78*<sup>-/-</sup> mutants (Fig. 3a-c). Nor did we detect ultrastructural defects: mutant glomus cells still contained the normal large dense core vesicles all along the plasma membrane and small clear core vesicles at synapses<sup>2</sup> (Fig. 3d-g). Thus, glomus cells are present in normal numbers and appear structurally intact in *Olfir78*<sup>-/-</sup> mutants.



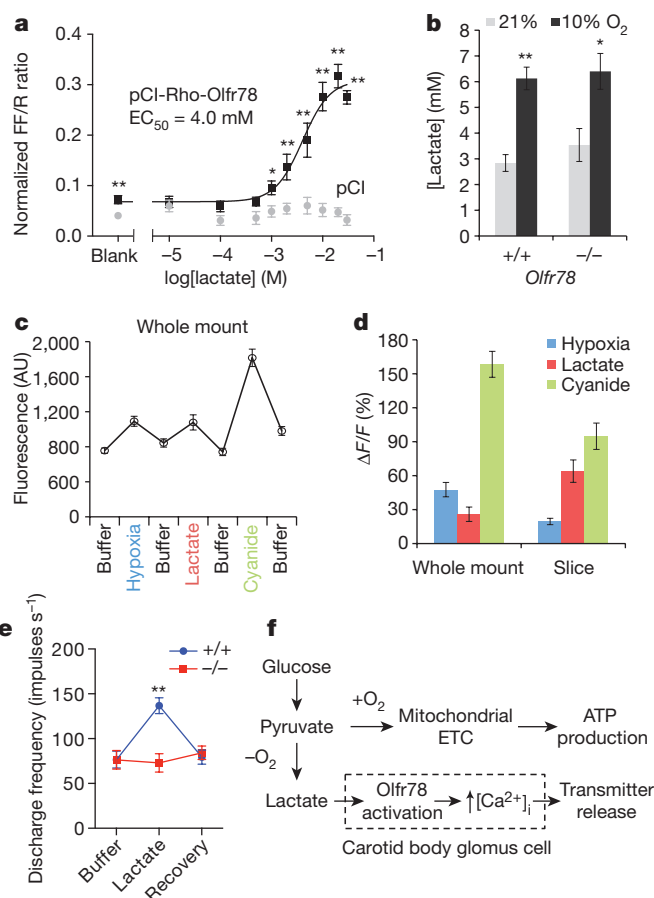
**Figure 2 | Ventilatory responses of *Olfir78* null mutants to hypoxia and hypercapnia.** Respiratory rate (RR), tidal volume (TV), and minute ventilation (MV) (minute ventilation = respiratory rate  $\times$  tidal volume) by whole body plethysmography of unrestrained, unanaesthetized *Olfir78*<sup>+/+</sup> and *Olfir78*<sup>-/-</sup> littermates exposed to hypoxia (**a, b**) or hypercapnia (**c, d**). **a, b**, Respiratory rate in hypoxia (**a**) and hypoxic response (**b**) as percentage change in hypoxia (10% O<sub>2</sub>) versus control (21% O<sub>2</sub>). Sample size  $n = 9$  ( $+/+$ ), 8 ( $-/-$ ) animals. **c, d**, Respiratory rate in hypercapnia (**c**) and hypercapnic response (**d**) as percentage change in hypercapnia (5% CO<sub>2</sub>) versus control (0% CO<sub>2</sub>). Sample size  $n = 4$  ( $+/+$ ), 5 ( $-/-$ ) animals. Data as mean  $\pm$  s.e.m. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  by paired  $t$ -test (**a, c**) or unpaired  $t$ -test (**b, d**).





To assess carotid body oxygen sensing, we performed extracellular recordings of the carotid sinus nerve (Extended Data Fig. 1b, c), a standard assay of carotid body neurosensory activity. We found that carotid sinus nerves from *Olfir78*<sup>-/-</sup> mutants had similar baseline discharge frequencies as *Olfir78*<sup>+/+</sup> and *Olfir78*<sup>+/-</sup> controls, demonstrating that nerve activity is intact in *Olfir78*<sup>-/-</sup> mutants. However, in hypoxia (PO<sub>2</sub> = 60–80 mmHg), control nerve activity increased substantially whereas *Olfir78*<sup>-/-</sup> mutant nerve activity showed little response (Fig. 3h and Extended Data Fig. 6a–h). By contrast, carotid sinus nerve activation by low pH was intact in *Olfir78*<sup>-/-</sup> mutants (Fig. 3i and Extended Data Fig. 6i, j). We conclude that *Olfir78* mutants have a specific defect in oxygen sensing in the carotid body.

Previous studies demonstrating the robust response of carotid body and breathing to cyanide and other electron transport inhibitors suggest that carotid body oxygen sensing may be mediated by a sensor that detects changes in metabolism<sup>2</sup> (Extended Data Fig. 1d, e). Interestingly, two short-chain fatty acids, acetate and propionate, activate *Olfir78* and its human orthologue OR51E2 expressed in HEK293T cells, with half maximal effective concentration (EC<sub>50</sub>) values of 1–3 mM<sup>14,19</sup>. However, blood concentrations of acetate and propionate in rodents and humans are only 0.1–0.3 mM and 4–25  $\mu$ M, respectively<sup>20,21</sup>, and change little in hypoxia relative to the *Olfir78* activation



curve<sup>22</sup>. Thus, we sought a ligand for *Olfir78* that is present in blood and tissue and effective at physiologically relevant levels. One appealing candidate that is chemically similar to acetate and propionate but more abundant *in vivo* is lactate, which is found in blood and tissue at low millimolar concentrations and rapidly increases in hypoxia (Extended Data Fig. 1e). Using a heterologous expression assay, we found that lactate activated *Olfir78* in a dose-dependent manner with an EC<sub>50</sub> of 4.0 mM (Fig. 4a and Extended Data Fig. 7a–g). Chloride ion over the same range of concentrations

curve<sup>22</sup>. Thus, we sought a ligand for *Olfir78* that is present in blood and tissue and effective at physiologically relevant levels.

One appealing candidate that is chemically similar to acetate and propionate but more abundant *in vivo* is lactate, which is found in blood and tissue at low millimolar concentrations and rapidly increases in hypoxia (Extended Data Fig. 1e). Using a heterologous expression assay, we found that lactate activated *Olfir78* in a dose-dependent manner with an EC<sub>50</sub> of 4.0 mM (Fig. 4a and Extended Data Fig. 7a–g). Chloride ion over the same range of concentrations

and osmolarity had no effect, whereas propionate and acetate stimulated Olf78 with EC<sub>50</sub> values similar to previous findings<sup>14</sup> (Extended Data Fig. 7d–g). Because lactate concentrations in blood, tissue, and tissue interstitium are 1–5 mM (Extended Data Fig. 1e), the observed EC<sub>50</sub> value of 4.0 mM for Olf78 renders it highly sensitive to small changes in lactate in the physiological range. Indeed, hypoxia and mitochondrial poisons such as cyanide elevate plasma and tissue lactate concentrations rapidly in this range (Extended Data Fig. 1e). We observed that arterial blood lactate increased from 3 mM to 6 mM within 3–5 min of hypoxia both in control and in *Olf78*<sup>−/−</sup> mutant animals (Fig. 4b and Extended Data Fig. 7h). Thus, lactate activates Olf78 in a physiologically relevant range.

Mitochondrial poisons trigger carotid body glomus cell activity<sup>23</sup>, and acute lactate application depolarizes glomus cells, stimulates carotid sinus nerve activity, and induces hyperventilation<sup>3,4,24,25</sup>. To determine whether lactate can directly activate glomus cells, we performed functional imaging experiments by expressing the calcium indicator GCaMP3 in glomus cells (Extended Data Fig. 8a–e). Both in whole carotid bodies and in slice preparations, we found that lactate induced calcium transients in glomus cells, as did hypoxia or addition of cyanide to inhibit mitochondrial electron transport chain and block oxygen consumption (Fig. 4c, d and Extended Data Fig. 8f–i). The response to lactate was stronger in slices than in intact carotid bodies, perhaps because glomus cells in slices have more direct exposure to lactate in the superfusate (Fig. 4d). We conclude that lactate can acutely activate glomus cells, much like hypoxia and cyanide. Interestingly, both in these experiments and in carotid sinus nerve recordings (ref. 25 and see below), carotid body activation by lactate was observed in hyperoxia, suggesting that lactate can stimulate carotid body sensory activity in the absence of other hypoxic signals.

To determine whether carotid body activation by lactate requires Olf78, we examined the effect of lactate on carotid sinus nerve activity in *Olf78*<sup>−/−</sup> mutants. Although lactate increased carotid sinus nerve activity in preparations from wild-type *Olf78*<sup>+/+</sup> animals as expected<sup>25</sup>, there was little response to lactate in *Olf78*<sup>−/−</sup> mutant nerves (Fig. 4e and Extended Data Fig. 6i, j). Similarly, acetate and propionate, two other Olf78 ligands that can also stimulate carotid sinus nerve activity<sup>14,26</sup>, had little effect in *Olf78*<sup>−/−</sup> mutants (Extended Data Fig. 6i, j). We conclude that carotid body activation by lactate and two other Olf78 ligands is mediated by Olf78.

Our results support a model in which decreased blood oxygen is sensed by the carotid body through an increase in production and secretion of lactate, which binds to Olf78 on glomus cells and induces calcium transients that increase signalling to afferent nerves to stimulate breathing (Fig. 4f). In the model, changes in blood oxygen are not detected directly by glomus cells, but indirectly through a metabolite (lactate) whose production is regulated by oxygen availability. This explains why drugs and mutations that inhibit the mitochondrial electron transport chain, preventing oxygen utilization and causing lactate build up, mimic the effect of hypoxia on carotid body activity and breathing, and supports the mitochondrial hypothesis of carotid body oxygen sensing (Extended Data Fig. 1d, e). Thus, the Olf78 pathway measures a metabolic state that integrates oxygen availability and demand and serves as a sentinel that signals, and attempts to stave off, an impending oxygen crisis, whereas the HIF-1 pathway senses oxygen directly (through prolyl hydroxylases that use oxygen to modify HIF-1 stability) and operates later and more broadly to deal with the crisis<sup>1</sup>.

What is the source of the lactate that activates Olf78? Lactate is produced by all cells in the body when oxygen declines: the blockage of mitochondrial electron transport leads to accumulation of upstream metabolites such as pyruvate, which is rapidly converted to lactate and then effluxed from cells (Extended Data Fig. 1e). Upon hypoxia exposure, lactate can almost double in blood within minutes (Fig. 4b and Extended Data Fig. 7h), and it accumulates in blood when inspired oxygen drops to levels that can activate carotid body

signalling and hyperventilation<sup>27</sup>. Besides blood, another potential source of lactate is the carotid body itself, as tissue lactate levels also increase rapidly in hypoxia, doubling within 30 s in some tissues<sup>5,6</sup>. Mitochondria of carotid body cells are highly sensitive to hypoxia compared with other tissues (Extended Data Fig. 1d), so when blood oxygen levels decline, carotid body cells should be among the first to produce lactate, ideal for Olf78 sentinel function. Because lactate is transported out of cells with H<sup>+</sup>, glomus cells would be exposed to extracellular increases in both lactate and H<sup>+</sup>, which could activate acid-sensitive channels (ASICs, TASKs) synergistically with Olf78 to stimulate glomus cells<sup>2,28,29</sup>. Lactate/Olf78 signalling may act with H<sup>+</sup> and perhaps other signals and pathways to promote the full carotid body response to hypoxia, explaining the small residual response to hypoxia detected in *Olf78*<sup>−/−</sup> mutants (Fig. 3h and Extended Data Fig. 6a–h).

In addition to the carotid body, Olf78 is expressed in several other oxygen-responsive tissues such as heart and lung<sup>9,13,14</sup>, and it is required for maintaining normal blood pressure<sup>14</sup>. We speculate that lactate and Olf78 serve as a general signal and sensor of hypoxia to control physiological responses. Nevertheless, some acute responses to hypoxia, such as reduced locomotion, regular breathing, and metabolic depression, are independent of Olf78. It may be valuable to develop synthetic agonists and antagonists of Olf78 for therapeutic control of breathing<sup>30</sup> and other responses it controls.

Genomic studies detect ectopic expression of other ORs in addition to Olf78 and Olf558, and some of these appear to be functional<sup>9,10,14</sup>. Downstream signalling in the carotid body may differ from that in olfaction (Extended Data Table 2), and it will be important to elucidate the full Olf78 signal transduction cascade in the carotid body and its integration with other pathways activated by hypoxia and other sensory stimuli. Although ORs were first identified for their role in smell, they may be involved in myriad chemosensory pathways detecting endogenous and exogenous ligands throughout the body.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 9 July; accepted 28 August 2015.**

1. Semenza, G. L. Hypoxia-inducible factors in physiology and medicine. *Cell* **148**, 399–408 (2012).
2. Kumar, P. & Prabhakar, N. R. Peripheral chemoreceptors: function and plasticity of the carotid body. *Compr. Physiol.* **2**, 141–219 (2012).
3. Lee, L. Y., Morton, R. F. & Lundberg, J. M. Pulmonary chemoreflexes elicited by intravenous injection of lactic acid in anesthetized rats. *J. Appl. Physiol.* **81**, 2349–2357 (1996).
4. Hardarson, T., Skarphedinsson, J. O. & Sveinsson, T. Importance of the lactate anion in control of breathing. *J. Appl. Physiol.* **84**, 411–416 (1998).
5. Kirsch, J. R. & D'Alecy, L. G. Role of tissue lactate and substrate availability in 1,3-butanediol-enhanced hypoxic survival in the mouse. *Stroke* **14**, 971–976 (1983).
6. Marina, N. et al. Brainstem hypoxia contributes to the development of hypertension in the spontaneously hypertensive rat. *Hypertension* **65**, 775–783 (2015).
7. Ganfornina, M. D. et al. Comparative gene expression profile of mouse carotid body and adrenal medulla under physiological hypoxia. *J. Physiol. (Lond.)* **566**, 491–503 (2005).
8. Fleischer, J., Breer, H. & Strotmann, J. Mammalian olfactory receptors. *Front. Cell. Neurosci.* **3**, 9 (2009).
9. Flegel, C., Mantiatis, S., Osthold, S., Hatt, H. & Gisselmann, G. Expression profile of ectopic olfactory receptors determined by deep sequencing. *PLoS One* **8**, e55368 (2013).
10. Kang, N. & Koo, J. Olfactory receptors in non-chemosensory tissues. *BMB Rep.* **45**, 612–622 (2012).
11. Conzelmann, S. et al. A novel brain receptor is expressed in a distinct population of olfactory sensory neurons. *Eur. J. Neurosci.* **12**, 3926–3934 (2000).
12. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
13. Weber, M., Pehl, U., Breer, H. & Strotmann, J. Olfactory receptor expressed in ganglia of the autonomic nervous system. *J. Neurosci. Res.* **68**, 176–184 (2002).
14. Pluznick, J. L. et al. Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. *Proc. Natl Acad. Sci. USA* **110**, 4410–4415 (2013).

15. Matsuura, S. Chemoreceptor properties of glomus tissue found in the carotid region of the cat. *J. Physiol. (Lond.)* **235**, 57–73 (1973).
16. Bozza, T. *et al.* Mapping of class I and class II odorant receptors to glomerular domains by two distinct types of olfactory sensory neurons in the mouse. *Neuron* **61**, 220–233 (2009).
17. Gautier, H. Interactions among metabolic rate, hypoxia, and control of breathing. *J. Appl. Physiol.* **81**, 521–527 (1996).
18. Shirahata, M., Kostuk, E. W. & Pichard, L. E. Carotid chemoreceptor development in mice. *Respir. Physiol. Neurobiol.* **185**, 20–29 (2013).
19. Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. D. Odor coding by a mammalian receptor repertoire. *Sci. Signal.* **2**, ra9 (2009).
20. Rémésy, C., Demigné, C. & Chartier, F. Origin and utilization of volatile fatty acids in the rat. *Reprod. Nutr. Dev.* **20** (Suppl. 4B), 1339–1349 (1980).
21. Wolever, T. M., Josse, R. G., Leiter, L. A. & Chiasson, J. L. Time of day and glucose tolerance status affect serum short-chain fatty acid concentrations in humans. *Metabolism* **46**, 805–811 (1997).
22. Knowles, S. E., Jarrett, I. G., Filsell, O. H. & Ballard, F. J. Production and utilization of acetate in mammals. *Biochem. J.* **142**, 401–411 (1974).
23. Buckler, K. J. & Turner, P. J. Oxygen sensitivity of mitochondrial function in rat arterial chemoreceptor cells. *J. Physiol. (Lond.)* **591**, 3549–3563 (2013).
24. Monti-Bloch, L., Abudara, V. & Eyzaguirre, C. Electrical communication between glomus cells of the rat carotid body. *Brain Res.* **622**, 119–131 (1993).
25. Pokorski, M. & Lahiri, S. Aortic and carotid chemoreceptor responses to metabolic acidosis in the cat. *Am. J. Physiol.* **244**, R652–R658 (1983).
26. Rigual, R., López-López, J. R. & Gonzalez, C. Release of dopamine and chemoreceptor discharge induced by low pH and high  $P_{CO_2}$  stimulation of the cat carotid body. *J. Physiol. (Lond.)* **433**, 519–531 (1991).
27. Huckabee, W. E. Relationships of pyruvate and lactate during anaerobic metabolism. III. Effect of breathing low-oxygen gases. *J. Clin. Invest.* **37**, 264–271 (1958).
28. Tan, Z. Y. *et al.* Acid-sensing ion channels contribute to transduction of extracellular acidosis in rat carotid body glomus cells. *Circ. Res.* **101**, 1009–1019 (2007).
29. Buckler, K. J. TASK channels in arterial chemoreceptors and their role in oxygen and acid sensing. *Pflugers Arch.* **467**, 1013–1025 (2015).
30. van der Schier, R., Roozkrans, M., van Velzen, M., Dahan, A. & Niesters, M. Opioid-induced respiratory depression: reversal by non-opioid drugs. *F1000Prime Rep.* **6**, 79 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank G. Fish for technical assistance; D. Riordan for assistance with RNA sequencing and analysis; L. He, B. Dinger, and S. Fidone for instruction on carotid body dissection; A. Gourine for instruction on carotid sinus nerve recordings; H. Matsunami for plasmids and advice about OR activity assays; A. Olson for assistance with two-photon imaging; R. Yu and K. Deisseroth for mouse strains; D. Cornfield and D. Bernstein for equipment; and P. Harbury and members of our laboratory for discussions and comments on the manuscript. This work was supported by Stanford University Dean's Postdoctoral Fellowship, National Institutes of Health (NIH) K12 RFA-HL-07-004 Career Development Program, Helen Hay Whitney Postdoctoral Fellowship, and NIH Pediatric Research Loan Repayment Program (A.J.C.), Howard Hughes Medical Institute Gilliam Fellowship (F.E.O.), NIH MH065541 and Harold and Leila Y. Mathers Charitable Foundation (D.V.M.), NIH NS069375 (Stanford Neuroscience Microscopy Service), and the Howard Hughes Medical Institute (M.A.K.).

**Author Contributions** A.J.C. conducted all the experiments except for luciferase assays, which were performed with F.E.O., and animal surgeries for blood gases, which were performed by J.R. Electrophysiology experiments were performed by A.J.C. in the laboratory of D.V.M., who participated in design, conduct, and analysis of these experiments. A.J.C. and M.A.K. conceived the experiments, analysed the data, and wrote the manuscript.

**Author Information** Microarray and RNA sequencing data have been deposited in the Gene Expression Omnibus database under accession number GSE72166. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.K. ([krasnow@stanford.edu](mailto:krasnow@stanford.edu)).



## METHODS

**Animals.** All experiments with animals were approved by the Institutional Animal Care and Use Committee at the Stanford University School of Medicine.

C57/BL6, stock 027 (Charles River) was used as wild type for microarrays and RNA sequencing. Other mouse strains used were as follows. *Olf78* knock-in reporter: MOL2.3-*IGITL*, shared by R. Yu<sup>11</sup>; *Olf78* knock-in mutant/reporter: B6;129P2-*Olf78*<sup>tm1Mom</sup>/MomJ, stock 006722 (JAX)<sup>16</sup>; *Olf78* mutant/reporter: B6129S5-*Olf78*<sup>tm1Lex</sup>, stock TF0586 (Taconic); Th-Cre driver: B6.FVB(Cg)-*Tg(Th-cre)*<sup>Fl172Gsat</sup>/Mmucd, stock 031029-UCD (MMRRC)<sup>31</sup>; Th-Cre driver: *Th<sup>tm1(cre)</sup>Te*, shared by K. Deisseroth<sup>32</sup>; ROSA-GCaMP3: B6;129S-Gt(ROSA)26Sor<sup>tm38(CAG-GCaMP3)Hze</sup>/J, stock 014538 (JAX)<sup>33</sup>; ROSA-tdTomato: B6;129S6-Gt(ROSA)26Sor<sup>tm9(CAG-tdTomato)Hze</sup>/J, stock 007905 (JAX)<sup>34</sup>.

Adult animals were used in all experiments, unless indicated otherwise. To control for sex differences, only female animals were used in physiology and behavioural experiments. Randomization and blinding were not used, in part because the *Olf78* mutant allele is genetically linked to a coat colour variant. All data include animals from multiple litters.

**Carotid body and adrenal medulla RNA purification.** Adult C57/BL6 animals were anaesthetized with isoflurane and decapitated, and carotid bifurcations and adrenal glands were dissected immediately and transferred to RNeasy lysis solution (Life Technologies) on ice. Carotid bodies and adrenal medullas were finely dissected from these tissues. From each animal, one or two carotid bodies and one adrenal medulla were obtained and stored in RNeasy lysis solution at 4 °C for up to 2 days. For each RNA purification, 18 carotid bodies and 10 adrenal medullas from 10 animals were pooled and processed using the RNeasy Micro Kit (Qiagen). Tissue pieces were disrupted in a guanidine-isothiocyanate lysis buffer (Buffer RLT, Qiagen) using a glass tissue grinder (Corning), homogenized using a 20 gauge needle and syringe, and purified by silica-membrane columns. RNA quality was assessed by electrophoresis on a Bioanalyzer using the RNA 6000 Pico Kit (Agilent). The average RNA integrity numbers for carotid body and adrenal medulla samples were 7.2 and 9.0, respectively.

**Microarrays.** Total RNA (>30 ng per sample) was processed using the 3' IVT Express Kit (Affymetrix) to make biotinylated amplified RNA (aRNA) by complementary DNA (cDNA) synthesis and *in vitro* transcription. aRNA was fragmented and hybridized to the GeneChip Mouse Genome 430 2.0 Array (Affymetrix) containing 45,000 probe sets targeting >34,000 mouse genes. aRNA synthesis, hybridization, and scanning were performed by the Stanford PAN Facility. Analysis of microarray data used Expression Console and Transcriptome Analysis Console software (Affymetrix).

**RNA sequencing.** Using the Amino Alkyl MessageAmp II aRNA Amplification Kit (Ambion), unlabelled aRNA was generated from total RNA (30–50 ng per sample) by an engineered M-MLV reverse transcriptase to make cDNA followed by *in vitro* transcription. aRNA was fragmented by RNA Fragmentation Reagents (Ambion), a buffered zinc solution, for 1.5 min at 70 °C. First- and second-strand cDNA synthesis, end repair, 3'-dA tail addition, and adaptor ligation were performed using the standard protocol from Illumina with adaptor oligonucleotides from Illumina, First Strand Buffer, SuperScript III reverse transcriptase, and Second Strand Buffer from Invitrogen, and RNaseH, DNA polymerase I, T4 DNA polymerase, Klenow DNA polymerase, T4 polynucleotide kinase, Klenow fragment (3'-5' exo-), and T4 DNA ligase from New England Biolabs. Modified cDNA libraries were resolved by electrophoresis in 2% low melting temperature agarose (Lonza) gels. For each sample, a region of the lane corresponding to ~200 base pairs was excised and purified by the QIAquick Gel Extraction Kit (Qiagen) using silica-membrane columns. Modified cDNA libraries were further amplified by PCR for 20 cycles using Phusion DNA polymerase (New England Biolabs). cDNA concentration and size were determined by electrophoresis using the High Sensitivity DNA Kit on the Bioanalyzer (Agilent), and samples were diluted to 4 pM for sequencing.

DNA clusters were generated using the Cluster Generation Kit according to manufacturer instructions (Illumina). Samples were then sequenced on the Illumina Genome Analyzer II using the 36-Cycle SBS Reagent Kit v2 (Illumina) run for 38 cycles. Each cDNA library was run in one lane, and data presented in this study are from two separate runs.

Sequences were aligned to the RefSeq database using Bowtie 0.9.8, allowing up to 4 mismatches in the first 32 bases for a sequence to be assigned to a specific gene ID. Reads that mapped to multiple isoforms of a gene were randomly assigned to one isoform, and counts for multiple messenger RNA (mRNA) isoforms for the same gene were combined for analysis. The number of aligned reads per 10<sup>7</sup> aligned reads was calculated after adding one read to every gene and sample to avoid dividing by 0 when calculating ratios between adrenal medulla and carotid body frequencies (Fig. 1a and Extended Data Fig. 2a). RPKM values were calculated by using the length of the mRNA, or the longest mRNA isoform

for genes that have multiple isoforms, in RefSeq (Fig. 1b and Extended Data Fig. 2b).

**X-gal staining.** Whole mount carotid bifurcations and adrenal glands were harvested, cleaned, and fixed in 4% paraformaldehyde (PFA)/PBS (pH 7.4) for 10 min at 22 °C. After washing with PBS, tissue was transferred to a solution of X-gal (1 mg ml<sup>-1</sup>), potassium ferricyanide (5 mM), potassium ferrocyanide (5 mM), magnesium chloride (2 mM), and NP-40 (0.02%) in PBS and incubated overnight at 37 °C. Samples were visualized on a Leica MZ12 stereomicroscope. Representative data reflect tissue samples from ten *Olf78*-GFP-*taulacZ*<sup>Tg/Tg</sup> (Fig. 1c–f, i–k) and three *Olf78*<sup>lacZ/lacZ</sup> (Extended Data Fig. 3e) animals.

For carotid bifurcation samples shown in section (n = 3 animals), tissue was fixed in 4% PFA/PBS for 10 min at 22 °C and equilibrated in 30% sucrose overnight at 4 °C. Tissue was then embedded in optimum cutting temperature compound (O.C.T., TissueTek) and stored at –80 °C. Sections were cut at 10 µm using a Leica CM3050S cryostat. X-gal solution was added onto sections on slides and incubated overnight at 37 °C. Slides were mounted in Mowiol 4–88 (Polysciences) with 1,4-diazabicyclo[2.2.2]octane (DABCO, 25 mg ml<sup>-1</sup>, Sigma-Aldrich) or Permount (45% polymer of α-pinene, β-pinene, dipentene, β-phellandrene/55% toluene, Fisher) and visualized on a Zeiss Axiophot fluorescence microscope.

For adult brain tissue, two animals were perfused through the heart with PBS followed by 4% PFA/PBS using a syringe. Whole brains were dissected from the head, and fixed again for 30 min in 4% PFA/PBS at 4 °C. After equilibration in 30% sucrose overnight at 4 °C, samples were embedded in O.C.T. (TissueTek) and sectioned at 80 µm using a Leica CM3050S cryostat. Slides were then incubated with X-gal overnight at 22 °C and visualized on a Leica MZ12 stereomicroscope.

**Immunostaining.** Tissue was fixed in 4% PFA/PBS at 22 °C for 10 min and equilibrated in 30% sucrose overnight at 4 °C. Tissue was embedded in O.C.T. (TissueTek) and stored at –80 °C. Sections were cut at 10 µm using a Leica CM3050S cryostat and incubated with primary antibodies overnight at 4 °C. Primary antibodies were rabbit anti-TH (Abcam, ab112), chicken anti-GFP (Abcam, ab13970), rat anti-CD31 (BD Pharmingen, 553370), chicken anti-β-galactosidase (Abcam, ab9361), and mouse anti-smooth muscle actin (Sigma, A5228) used at 1:500. Mouse anti-smooth muscle actin antibody was directly conjugated to Cy5 NHS ester (GE Healthcare), and unbound dye was removed on a P-30 gel exclusion column (BioRad)<sup>35</sup>. Incubation with secondary antibodies was 45 min at 22 °C. Secondary antibodies were conjugated to either Alexa Fluor 488, Alexa Fluor 555 (Life Technologies), or DyLight 488 (Jackson ImmunoResearch). Staining with DAPI (1 ng ml<sup>-1</sup>, Life Technologies) was performed after incubation with secondary antibodies for 5 min at 22 °C. Sections were mounted in Mowiol 4–88 (Polysciences) with DABCO (25 mg ml<sup>-1</sup>, Sigma-Aldrich) and visualized on a Zeiss Axiophot fluorescence microscope. Tissue samples from three or more animals were stained for representative data shown (Fig. 1g, h, Fig. 3a, b, and Extended Data Fig. 3a–c, f–h).

**Electron microscopy.** Carotid bifurcations were harvested from adult animals and transferred to fixation solution (4% PFA and 2% glutaraldehyde in PBS) for 1 h at 22 °C. During fixation, excess tissue was trimmed away to retain the carotid body and carotid arteries. Samples were post-fixed with osmium tetroxide for 1.75 h at 4 °C, washed three times with cold double-distilled H<sub>2</sub>O, and incubated with 1% uranyl acetate overnight at 4 °C. The next day, samples were serially dehydrated in ethanol (50%, 70%, 100%, 100%) for 10 min per step and washed with propylene oxide for 15 min, all at 22 °C. Samples were then transferred to 1:1 propylene oxide:Epon (Electron Microscopy Services) for 1 h, 1:2 propylene oxide:Epon for 45 min, and 100% Epon, all at 22 °C. Once samples were embedded in 100% Epon, blocks were baked overnight at 65 °C.

To locate the carotid body in the embedded tissue, sections were cut at 2 µm using glass knives on a Leica Ultracut S microtome and stained with Toluidine Blue (Sigma-Aldrich) for visualization of tissue histology. Once the carotid body was reached, 17 nm sections were cut using a diamond blade for transmission electron microscopy. Sections were visualized on a JEOL TEM1230 transmission electron microscope equipped with a Gatan 967 slow-scan, cooled CCD camera. Two sections at different levels in the carotid body were examined for each sample. All sectioning and imaging procedures were performed at the Stanford Cell Sciences Imaging Facility-Electron Microscopy Core.

**Whole body plethysmography.** Unrestrained, unanaesthetized adult animals were transferred to a whole body plethysmograph (450 ml, Model PY4211, Buxco) connected to a MAXII preamplifier unit and computer running BioSystem XA software (Buxco). To reach stable baseline ventilation, animals were acclimatized to the chamber for more than 30 min in control gas conditions before exposure to hypoxia or hypercapnia. Three pulses of hypoxia or hypercapnia lasting 5 min each were performed with 10-min recovery periods in control conditions. Gas mixtures for control, hypoxia, and hypercapnia were 21% O<sub>2</sub>/79% N<sub>2</sub>, 10% O<sub>2</sub>/90% N<sub>2</sub>, and

5% CO<sub>2</sub>/21% O<sub>2</sub>/74% N<sub>2</sub>, respectively (Praxair). Flow rates were 1.5 l min<sup>-1</sup> during measurement periods and 11–12 l min<sup>-1</sup> during 1-min ramp periods between gas mixtures.

Ventilatory parameters were collected and calculated by BioSystem XA software (Buxco). Tidal volumes were calculated according to Drorbaugh and Fenn<sup>36</sup> with manual input of environmental conditions, such as room and chamber temperature, humidity, and barometric pressure. To enrich for measurements of regular breaths, criteria were set in the software to accept a breath if (1) inspiratory time was greater than 0.07 s, (2) expiratory time was less than 10 s, (3) calculated tidal volume was greater than 0.05 ml, and (4) volume balance between inspiration and expiration was less than 50%. Under these conditions, virtually all breaths were accepted during very regular breathing in hypoxia and hypercapnia. Fifteen breaths were averaged for each line of data, and lines of data for each period of control or stimulus were averaged, excluding lines that had more than two observed events of sniffing, grooming, or movement among the accepted breaths. Ventilatory parameters over all periods of control or stimulus were averaged for each animal and are presented in the figures. Most animals were tested twice within one week with good reproducibility, and measurements were averaged. Numbers of animals tested were comparable to other published work<sup>37–40</sup>. Although there was no formal randomization, different numbers, genotypes, and litters were tested in different orders on multiple days over several months. For the animals used in our study, body weight did not correlate with respiratory rate, tidal volume, or minute ventilation in wild-type animals, mutant animals, or all animals together (correlation coefficients  $0.001 \leq R^2 \leq 0.289$ ,  $0.996 \geq P \geq 0.293$ ).

**Blood gas and lactate measurements.** For blood testing under anaesthetized conditions, animals were transferred to individual cages on the morning of testing and allowed to acclimate for at least 4 h. This was designed to avoid repeated cage handling and removal of other animals from the same cage, procedures that have been shown to increase stress hormones and blood lactate levels in rodents<sup>41</sup>. Animals were quickly anaesthetized with 3% isoflurane in 100% O<sub>2</sub> in an acrylic container and maintained in 1.5–2% isoflurane in 21% O<sub>2</sub>/79% N<sub>2</sub> or 10% O<sub>2</sub>/90% N<sub>2</sub> (Praxair) at 2 l min<sup>-1</sup> through a nose cone. Body temperature was maintained at 37 °C using a heating pad with feedback temperature controller (Physitemp Instruments). The right carotid artery was surgically isolated and cut, and 200–250 µl of blood was collected using a heparinized syringe. An aliquot (~100 µl) of arterial blood was immediately transferred to a CG4+ cartridge for measurement of blood gases and lactate using an i-Stat Portable Clinical Analyzer (Abbott). Time from beginning of surgery to blood collection was ~3 min, during which the hypoxic ventilatory response was still robust under our conditions (data not shown). For some animals, arterial blood was also transferred to a test strip for lactate analysis using a Lactate Scout analyser (EKF Diagnostics). We found good correlation between lactate measurements of the same blood sample from anaesthetized animals using i-Stat and Lactate Scout analysers ( $n = 11$ ,  $R^2 = 0.92$ ,  $P < 0.0001$ ).

For blood lactate measurements in unanaesthetized conditions, animals were transferred to individual cages at least 1 day before the first day of testing. Animals in their housing cage were moved into a hypoxia control glove box set to 21% O<sub>2</sub> or 10% O<sub>2</sub> balanced by N<sub>2</sub> (Coy Laboratory Products). After 1 min in the airlock, the cage was moved into the glove box, and the lid was opened for another 3 min. Then the animal was transferred to a tail vein restrainer (Baintree Scientific), and the tail artery was punctured with a 27G½ needle. Blood was then directly transferred to a test strip for measurement of lactate using a Lactate Scout analyser. The total time of animals in the glove box before blood testing was 4–5 min. Owing to handling stress increasing blood lactate concentrations<sup>41</sup> and causing more variable blood lactate measurements in awake conditions, animals were kept in the same room and tested on four separate days for 2 days each of 21% O<sub>2</sub> or 10% O<sub>2</sub> exposure. Results for each animal and oxygen condition were then averaged. One *Olf1r78*<sup>+/+</sup> animal was excluded because of excessive handling stress due to long blood collection on 2 days.

**Body temperature measurements.** Unanaesthetized animals were transferred to a tail vein restrainer (Baintree Scientific), and body temperature was measured using a rectal temperature probe and animal temperature controller (Physitemp Instruments) in room air (21% O<sub>2</sub>) or in 10% O<sub>2</sub> in a hypoxia control glove box (Coy Laboratory Products). Data were collected in 21% O<sub>2</sub> 1 h before transfer to the hypoxia control glove box and at 2 and 5 min in 10% O<sub>2</sub> inside the glove box. An airlock was used to transfer the animal into the glove box for a ramp time of 1 min from 21% O<sub>2</sub> to 10% O<sub>2</sub>.

**Metabolic measurements.** Unrestrained, unanaesthetized animals were transferred to the same chamber used for plethysmography, which was sealed to allow only airflow in from the side port and out from a bottom port on the opposite side. Metabolic measurements were collected using an OxyMax open circuit indirect

calorimeter (Columbus Instruments) with an electrochemical oxygen sensor modified to measure two ranges around 21% O<sub>2</sub> and 10% O<sub>2</sub>. Flow was set to 0.6 l min<sup>-1</sup> from gas mixtures of 21% O<sub>2</sub>/79% N<sub>2</sub> and 10% O<sub>2</sub>/90% N<sub>2</sub> (Praxair). Measurements were taken every 30 s. For 21% O<sub>2</sub>, animals were allowed to acclimate to the chamber for 10 min, and data are shown for 10–15 min after the start of measurements. For 10% O<sub>2</sub>, animals became calm more quickly, and data are shown for 5–10 min after the start of measurements, a duration we found necessary for the system to stabilize to 10% O<sub>2</sub> after opening the chamber.

**Measurements of oxygen in perfusion.** Measurements of oxygen concentrations of the perfusion solution in the recording chamber were performed using a Clark-style oxygen electrode (Unisense). Because voltage readings for the sensor were observed to be highly temperature-dependent, the sensor was calibrated at the temperature of the relevant protocol, which was 33–34 °C for electrophysiology and 22 °C for calcium imaging.

**Carotid sinus nerve recordings.** Animals were terminally anaesthetized with isoflurane, perfused through the heart with ice-cold artificial cerebrospinal fluid (ACSF, pH 7.4) composed of 119 mM NaCl, 5 mM KCl, 2.5 mM CaCl<sub>2</sub>, 1.3 mM MgSO<sub>4</sub>, 1 mM NaH<sub>2</sub>PO<sub>4</sub>, 26.2 mM NaHCO<sub>3</sub>, and 11 mM glucose previously bubbled with 95% O<sub>2</sub>/5% CO<sub>2</sub> (Praxair), and decapitated. Both carotid bifurcations were then dissected in ice-cold ACSF and transferred to a recording chamber (3 ml), where they were superfused with ACSF continuously bubbled with 95% O<sub>2</sub>/5% CO<sub>2</sub> (Praxair) at a flow rate of 13.3 ml min<sup>-1</sup> by gravity and maintained at 33–34 °C. The carotid sinus nerve was carefully exposed and cut near the point where it branches from the glossopharyngeal nerve. The cut end was pulled into a tightly fitting glass suction micropipette<sup>38,40</sup>, and voltage was recorded relative to a reference in the bath using an Axoclamp 2A electrometer (Molecular Devices) in Bridge mode. The voltage signal was amplified 1,000 times (10 times on the Axoclamp and 100 times on a Brownlee Precision Model 440 instrumentation amplifier), filtered (0.2–3 kHz), and digitized at 10 kHz on a National Instruments MIO15E-2 analogue-to-digital converter. Data were stored and displayed using in-house software written in LabView (National Instruments).

If spikes were not observed at baseline for the first carotid sinus nerve, the second carotid bifurcation was dissected and recorded. One-second sweeps were acquired continuously through the entire time course, and only one carotid sinus nerve recording per animal was included in the data presented. Hypoxia stimulus was delivered by changing the gas bubbling the ACSF to 95% N<sub>2</sub>/5% CO<sub>2</sub> (Praxair) for 8 min. Under these conditions, PO<sub>2</sub> levels in the recording chamber started at 625 mmHg and decreased to a low of 60 mmHg by 9 min after the start of bubbling with 95% N<sub>2</sub>/5% CO<sub>2</sub> (Fig. 3h). Solutions of lactate, acetate, and propionate (30 mM, pH 7.4) were made by equimolar substitution of NaCl in ACSF with sodium salts of L-lactate, acetate, and propionate, respectively. Low pH solution (pH 7.0) was made by lowering the NaHCO<sub>3</sub> in ACSF from 26.2 mM to 11.9 mM with an equimolar increase in NaCl<sup>38</sup>. These solutions were continuously bubbled with 95% O<sub>2</sub>/5% CO<sub>2</sub> (Praxair) at the reservoir, maintaining the oxygen concentration of the solution in the chamber at PO<sub>2</sub> = 625 mmHg. All preparations were stimulated with bolus injections of 25–50 µl sodium cyanide (20 mM) at the end of the experiment to confirm that the nerve was active.

Recordings were analysed offline using Spike2 software (Cambridge Electronic Design). To measure action potential frequency, we analysed a 1-s period of data every minute. A single threshold was used for spike determination and set empirically for each time course by moving the threshold through a range of values until the spike count stabilized through several intervals of 0.001 mV and then dropped off for data at time = 0 and 9 min (hypoxia) or all time points scored (acetate, propionate, lactate, and low pH). The lowest threshold value in the stable range was applied to all sweeps of a stimulus analysed for each recording. Because we noticed that spikes close together or with low amplitudes were often missed by the software, we also manually counted spikes for the same sweeps analysed by software. Two recordings were excluded owing to low signal to noise precluding accurate analysis. In two experiments, we also applied 7.5 µM tetrodotoxin (TTX) to block voltage-gated sodium ion channels during hypoxia exposure as an additional confirmation that events being scored were action potentials (Extended Data Fig. 6c, d).

**pCI-Rho-Olf1r78 plasmid construction and expression.** The pCI-Rho-Olf1r78 plasmid was made by PCR amplifying the *Olf1r78* coding region from a pCMV6-Olf1r78 plasmid (OriGene) using forward primer 5'-ATTGCCGAATTCATGATTCCTGCAACTTCACC-3' and reverse primer 5'-ATTGCCGCGGCCGCTACGTGTTTCCCCAGCTTCAA-3', adding EcoRI and NotI restriction sites. The *Olf1r78* PCR fragment was then digested with EcoRI and NotI and cloned into a pCI-Rho backbone cut from a pCI-Rho-Olf1r62 plasmid (a gift from H. Matsunami). Cell-surface expression of Rho-epitope-tagged *Olf1r78* protein in HEK293T cells was confirmed by immunostaining as described<sup>42</sup>. Cytoplasmic



GFP (co-transfection marker) was expressed from a TBC1D25::eGFP plasmid (a gift from S. Pfeffer).

**Luciferase assay.** HEK293T cells were grown and seeded into 96-well plates as described<sup>42</sup>. On the next day, cells were transfected with RTP1S,  $G_{\alpha 15}$ -olf (gifts from H. Matsunami), pCMV6-Ric8b (Origene), pCRE-Luc (Agilent), and pSV40-RL (Promega) plasmids and either pCI-Rho-Olf78 or pCI (Promega). The Rho tag on Olf78 and RTP1S were used to enhance localization of Olf78 to the plasma membrane.  $G_{\alpha 15}$ -olf and Ric8b were included as downstream effectors that couple to ORs to increase cAMP production upon OR activation. Two transcriptional luciferase reporters, one constitutive (*Renilla*, pSV40-RL) and one inducible by cAMP (firefly, pCRE-Luc), were transfected to report increased cAMP levels upon OR activation<sup>42</sup>.

Five hours after transfection, media was decanted and replaced with 50  $\mu$ l per well MEM without phenol red (Life Technologies). After 30 min, 25  $\mu$ l of chemicals were added to each well to achieve the indicated final concentrations, and cells were incubated for 2 h. Sodium salts of chloride, L-lactate, propionate, and acetate (Sigma) were used. The duration of transfection and stimulation was shortened because we observed that transfection of cells with pCI-Rho-Olf78 overnight caused a large increase in firefly luciferase activity in the absence of added chemicals (data not shown), suggesting that lactate or some other molecule released from cells and/or a component of the transfection mixture could stimulate Olf78 activity.

Reagents to detect firefly and *Renilla* luciferase activity (Dual-Glo Luciferase Assay System, Promega) were added at 20  $\mu$ l per well<sup>42</sup>. Luminescence was measured using an Infinite M1000 (Tecan) microplate reader and data acquired by Magellan Data Analysis Software (Tecan). Two readings were collected for each plate and luciferase reagent, and firefly:*Renilla* ratios were averaged. Data were from experiments conducted over 3 days. For dose–response curves in Fig. 4a and Extended Data Fig. 7d–f, ratios were normalized to the highest and lowest average values for a given condition across all plates on each day.

The HEK293T cell line (a gift from S. Pfeffer) was not authenticated or tested for mycoplasma contamination.

**Calcium imaging.** *Th-Cre*; *ROSA-GCaMP3* animals expressing GCaMP3 in glomus cells were generated using two *Th-Cre* drivers<sup>31–33</sup>. Both *Th-Cre* lines drove expression in glomus cells, as confirmed by *Th-Cre*; *ROSA-tdTomato* animals<sup>34</sup>, but the BAC-transgenic *Th-Cre* driver (MMRRC) required two copies of *ROSA-GCaMP3* reporter for robust expression. For whole mount preparations, carotid bifurcations were dissected from transgenic animals and transferred to 0.5% glucose/PBS bubbling 100% O<sub>2</sub> on ice. Surrounding tissue was removed to expose the carotid body attached to the carotid artery. The carotid body was incubated in a physiological buffer (115 mM NaCl, 5 mM KCl, 24 mM NaHCO<sub>3</sub>, 1 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>, 11 mM glucose<sup>43</sup>) at 37 °C in a tissue culture incubator with 5% CO<sub>2</sub> before transfer to the recording chamber for imaging.

For tissue slices, carotid bifurcations were dissected and transferred to a modified Tyrode's solution (148 mM NaCl, 2 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM HEPES, 10 mM glucose), pH 7.4 on ice<sup>44</sup>. Carotid bodies were then isolated and embedded in 3% low melting temperature agarose (Lonza) in a sample holder (Precisionary Instruments). Tissue slices were cut at 100  $\mu$ m using a Compresstome VF-200 (Precisionary Instruments). Samples were then transferred to culture medium composed of DMEM with 10% FBS, 1% penicillin/streptomycin, and Insulin-Transferrin-Selenium (Life Technologies) and incubated in a tissue culture incubator at 37 °C with 5% CO<sub>2</sub> for at least 24 h before calcium imaging according to established protocols<sup>44</sup>.

At baseline, the carotid body was superfused with physiological buffer bubbling 95% O<sub>2</sub>/5% CO<sub>2</sub> at 3.75 ml min<sup>−1</sup> using a Reglo analogue tubing pump (Ismatec), maintaining the oxygen concentration of the solution in the chamber at PO<sub>2</sub> = 600 mmHg. Hypoxia was generated by bubbling physiological buffer with 95% N<sub>2</sub>/5% CO<sub>2</sub>. Lactate solution (30 mM) was made by equimolar substitution of NaCl with sodium L-lactate. Lactate and cyanide solutions were bubbled with 95% O<sub>2</sub>/5% CO<sub>2</sub>. To switch between stimuli, the flow rate was increased to 7.5 ml min<sup>−1</sup> for 2 min.

The carotid body was imaged on a Prairie Ultima XY two-photon rig built around an Olympus BX-61W upright microscope at the Stanford Neuroscience Microscopy Service. Using a water immersion 60 $\times$  objective, Z-stacks at 2  $\mu$ m steps were collected at a resolution of 1,024 pixels  $\times$  1,024 pixels for 70–100  $\mu$ m of tissue. Two stacks were collected for hypoxia and lactate stimuli and intervening buffer recovery periods for whole mount samples. Images were analysed using ImageJ. Regions of interest corresponding to individual glomus cells were defined by the images with the strongest fluorescence. Average fluorescence intensities were calculated for each region of interest, and values were averaged for stimulus and buffer periods with more than one stack. Cells that had very high levels of fluorescence at the start of the experiment were excluded from our analysis of data from whole

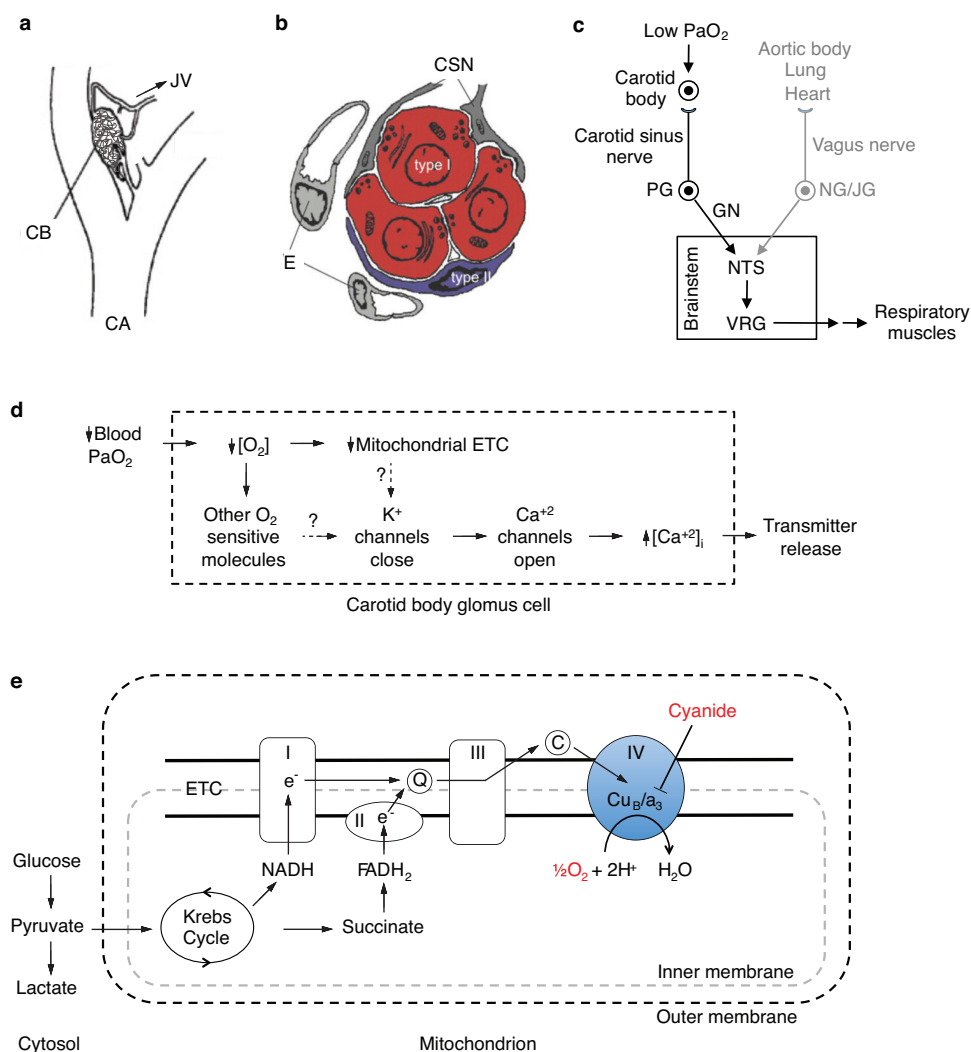
mount samples<sup>43</sup> because they showed dramatic declines in baseline fluorescence after hypoxia and lactate stimulation (data not shown). Data presented are from two samples from two different animals performed on separate days.

**Data analysis and statistics.** Data analysis and statistical tests were performed using Microsoft Excel and GraphPad software. GraphPad Prism 6 was used to fit dose–response curves using a variable slope model and to calculate EC<sub>50</sub> values. All data are biological replicates, and quantitative data with error bars are presented as mean  $\pm$  s.e.m. with the exception of Extended Data Fig. 7c, which is presented as percentage  $\pm$  standard error of percentage. Groups compared by parametric tests fit the assumption for normal distribution as determined by the Shapiro–Wilk test with the critical *W* value set at 5% significance level. All *t*-tests shown are two-sided, and variances of sample groups compared were similar. No statistical method was used to predetermine sample size.

- Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).
- Lindeberg, J. *et al.* Transgenic expression of Cre recombinase from the tyrosine hydroxylase locus. *Genesis* **40**, 67–73 (2004).
- Zariwala, H. A. *et al.* A Cre-dependent GCaMP3 reporter mouse for neuronal imaging in vivo. *J. Neurosci.* **32**, 3131–3141 (2012).
- Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
- Kumar, M. E. *et al.* Mesenchymal cells. Defining a mesenchymal progenitor niche at single-cell resolution. *Science* **346**, 1258810 (2014).
- Drorbaugh, J. E. & Fenn, W. O. A barometric method for measuring ventilation in newborn infants. *Pediatrics* **16**, 81–87 (1955).
- Peng, Y. J. *et al.* H<sub>2</sub>S mediates O<sub>2</sub> sensing in the carotid body. *Proc. Natl Acad. Sci. USA* **107**, 10719–10724 (2010).
- Trapp, S., Tucker, S. J. & Gourine, A. V. Respiratory responses to hypercapnia and hypoxia in mice with genetic ablation of Kir5.1 (*Kcnj16*). *Exp. Physiol.* **96**, 451–459 (2011).
- Kline, D. D., Peng, Y. J., Manalo, D. J., Semenza, G. L. & Prabhakar, N. R. Defective carotid body function and impaired ventilatory responses to chronic hypoxia in mice partially deficient for hypoxia-inducible factor 1  $\alpha$ . *Proc. Natl Acad. Sci. USA* **99**, 821–826 (2002).
- Kline, D. D. *et al.* Kv1.1 deletion augments the afferent hypoxic chemosensory pathway and respiration. *J. Neurosci.* **25**, 3389–3399 (2005).
- Balcombe, J. P., Barnard, N. D. & Sandusky, C. Laboratory routines cause animal stress. *Contemp. Top. Lab. Anim. Sci.* **43**, 42–51 (2004).
- Zhuang, H. & Matsunami, H. Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nature Protocols* **3**, 1402–1413 (2008).
- Piskuric, N. A. & Nurse, C. A. Effects of chemostimuli on [Ca<sup>2+</sup>]<sub>i</sub> responses of rat aortic body type I cells and endogenous local neurons: comparison with carotid body cells. *J. Physiol. (Lond.)* **590**, 2121–2135 (2012).
- Pardal, R., Ludewig, U., Garcia-Hirschfeld, J. & Lopez-Barneo, J. Secretory responses of intact glomus cells in thin slices of rat carotid body to hypoxia and tetraethylammonium. *Proc. Natl Acad. Sci. USA* **97**, 2361–2366 (2000).
- Pardal, R., Ortega-Sáenz, P., Durán, R. & López-Barneo, J. Glia-like stem cells sustain physiologic neurogenesis in the adult mammalian carotid body. *Cell* **131**, 364–377 (2007).
- Gourine, A. V. On the peripheral and central chemoreception and control of breathing: an emerging role of ATP. *J. Physiol. (Lond.)* **568**, 715–724 (2005).
- Anickhov, S. V. & Belen'ii, M. L. *Pharmacology of the Carotid Body Chemoreceptors* (Pergamon, 1963).
- López-Barneo, J., López-López, J. R., Ureña, J. & González, C. Chemotransduction in the carotid body: K<sup>+</sup> current modulated by PO<sub>2</sub> in type I chemoreceptor cells. *Science* **241**, 580–582 (1988).
- Buckler, K. J. A novel oxygen-sensitive potassium current in rat carotid body type I cells. *J. Physiol. (Lond.)* **498**, 649–662 (1997).
- Williams, S. E. *et al.* Hemoxxygenase-2 is an oxygen sensor for a calcium-sensitive potassium channel. *Science* **306**, 2093–2097 (2004).
- Wyatt, C. N. *et al.* AMP-activated protein kinase mediates carotid body excitation by hypoxia. *J. Biol. Chem.* **282**, 8092–8098 (2007).
- Streller, T., Huckstorf, C., Pfeiffer, C. & Acker, H. Unusual cytochrome a<sub>592</sub> with low PO<sub>2</sub> affinity correlates as putative oxygen sensor with rat carotid body chemoreceptor discharge. *FASEB J.* **16**, 1277–1279 (2002).
- Mills, E. & Jöbsis, F. F. Mitochondrial respiratory chain of carotid body and chemoreceptor response to changes in oxygen tension. *J. Neurophysiol.* **35**, 405–428 (1972).
- Wilson, D. F. *et al.* The primary oxygen sensor of the cat carotid body is cytochrome a<sub>3</sub> of the mitochondrial respiratory chain. *FEBS Lett.* **351**, 370–374 (1994).
- Quintana, A. *et al.* Fatal breathing dysfunction in a mouse model of Leigh syndrome. *J. Clin. Invest.* **122**, 2359–2368 (2012).
- Stettner, G. M., Viscomi, C., Zeviani, M., Wilichowski, E. & Dutschmann, M. Hypoxic and hypercapnic challenges unveil respiratory vulnerability of Surf1 knockout mice, an animal model of Leigh syndrome. *Mitochondrion* **11**, 413–420 (2011).

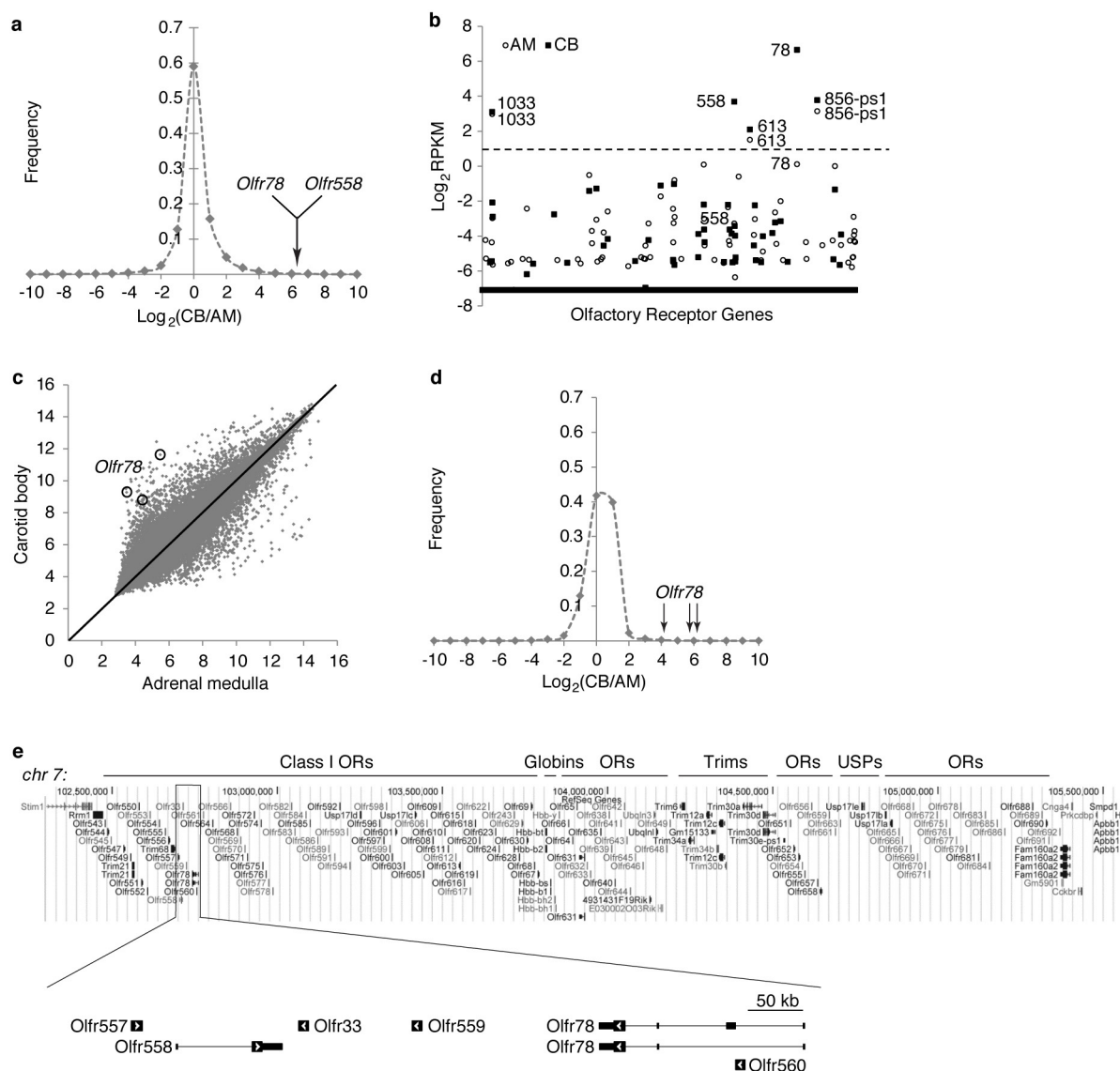


57. Pronicka, E. *et al.* Compulsory hyperventilation and hypocapnia of patients with Leigh syndrome associated with SURF1 gene mutations as a cause of low serum bicarbonates. *J. Inherit. Metab. Dis.* **24**, 707–714 (2001).
58. Piruat, J. I., Pintado, C. O., Ortega-Sáenz, P., Roche, M. & López-Barneo, J. The mitochondrial SDHD gene is required for early embryogenesis, and its partial deficiency results in persistent carotid body glomus cell activation with full responsiveness to hypoxia. *Mol. Cell. Biol.* **24**, 10933–10940 (2004).
59. Mulligan, E., Lahiri, S. & Storey, B. T. Carotid body O<sub>2</sub> chemoreception and mitochondrial oxidative phosphorylation. *J. Appl. Physiol.* **51**, 438–446 (1981).
60. Moore, S. J., Ho, I. K. & Hume, A. S. Severe hypoxia produced by concomitant intoxication with sublethal doses of carbon monoxide and cyanide. *Toxicol. Appl. Pharmacol.* **109**, 412–420 (1991).
61. Halestrap, A. P. Monocarboxylic acid transport. *Compr. Physiol.* **3**, 1611–1643 (2013).
62. Schwarzkopf, T. M., Horn, T., Lang, D. & Klein, J. Blood gases and energy metabolites in mouse blood before and after cerebral ischemia: the effects of anesthetics. *Exp. Biol. Med. (Maywood)* **238**, 84–89 (2013).
63. Rosdahl, H., Ungerstedt, U., Jorfeldt, L. & Henriksson, J. Interstitial glucose and lactate balance in human skeletal muscle and adipose tissue studied by microdialysis. *J. Physiol. (Lond.)* **471**, 637–657 (1993).
64. Balbir, A. *et al.* A search for genes that may confer divergent morphology and function in the carotid body between two strains of mice. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **292**, L704–L715 (2007).
65. Yokoyama, T., Misuzu, Y. Y. & Yamamoto, Y. Immunohistochemical localization of tryptophan hydroxylase and serotonin transporter in the carotid body of the rat. *Histochem. Cell Biol.* **140**, 147–155 (2013).
66. Izal-Azcárate, A. *et al.* Immunohistochemical characterization of the rat carotid body. *Respir. Physiol. Neurobiol.* **161**, 95–99 (2008).



**Extended Data Figure 1 | Model of oxygen sensing by the carotid body and the mitochondrion.** **a**, Anatomy and blood supply of the carotid body. The carotid body is located bilaterally at bifurcation of carotid artery (CA) in the neck. Its location can be variable as well as the source of its blood supply, which can come from branches of nearby internal and external carotid, occipital, pharyngeal arteries. Blood flows through fenestrated capillaries close to clusters of type I glomus cells and drains from carotid body into jugular vein (JV) on ventral side<sup>2</sup>. **b**, Cellular organization of carotid body. The carotid body is composed of several cell types, including type I glomus cells (red) that sense changes in blood oxygen and are organized in clusters, type II sustentacular cells (blue) that resemble neuroglia and surround glomus cell clusters, carotid sinus nerve (CSN) fibres that innervate glomus cells, and endothelial (E) and smooth muscle cells (not shown) that form the tortuous vasculature<sup>2</sup>. Panels **a** and **b** modified from Pardal, R., Ortega-Sáenz, P., Durán, R. & López-Barneo, J. Glia-like stem cells sustain physiologic neurogenesis in the adult mammalian carotid body, *Cell* **131**, 364–377 (2007) (ref. 45) with permission from Elsevier. **c**, Oxygen-sensing respiratory circuit. The primary chemoreceptor for blood oxygen is the carotid body. A decrease in  $\text{PaO}_2$  of arterial blood from normoxia (100 mmHg) to hypoxia (< 80 mmHg) stimulates glomus cells to signal the carotid sinus nerve, a branch of glossopharyngeal nerve (GN) with cell bodies in petrosal ganglion (PG). Axons of the GN terminate in nucleus tractus solitarius (NTS) in brainstem, a site of many converging afferent inputs<sup>2</sup>. The signal from NTS is transmitted to the ventral respiratory group (VRG) that includes the pre-Bötzinger complex, a region essential for respiratory rhythm generation. From VRG, neurons project to premotor and motor neurons that innervate respiratory muscles, such as diaphragm and intercostal muscles<sup>46</sup>. In addition to carotid body, vagus nerve

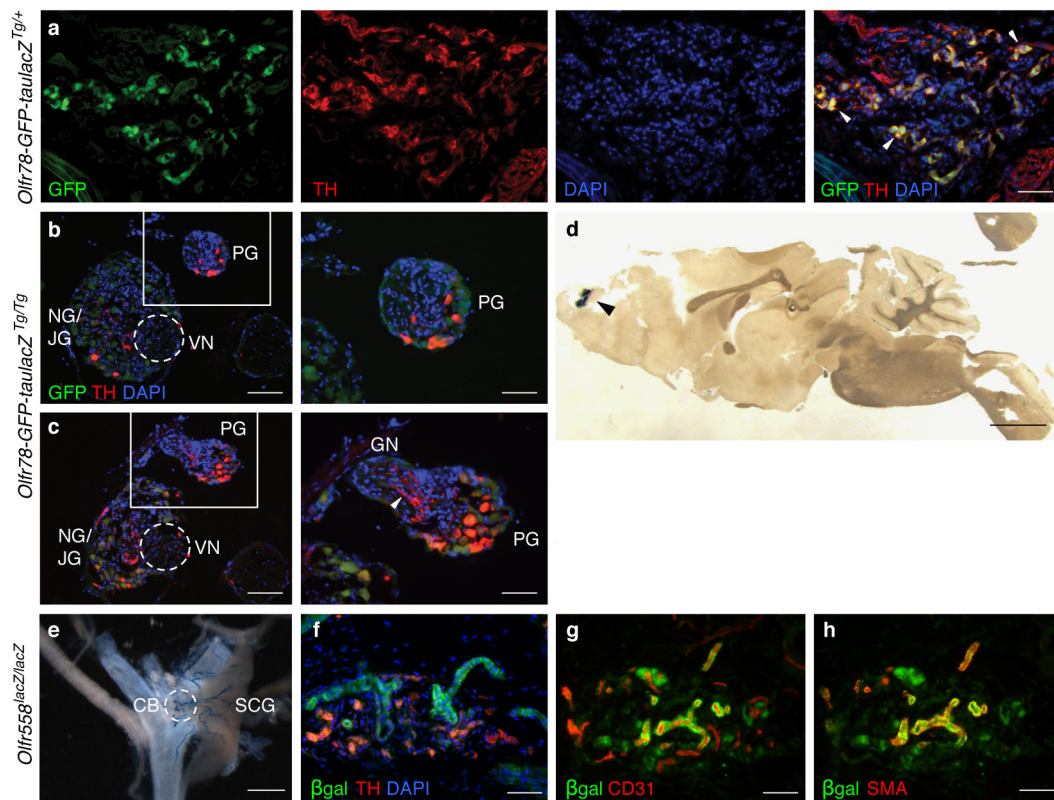
afferents can also contribute to respiratory behaviours under specialized conditions<sup>39</sup>. The vagus nerve innervates heart, lung, and oxygen-sensitive cells of the aortic body<sup>43</sup>. **d**, A current model of acute oxygen sensing by carotid body. A decrease in  $\text{PaO}_2$  in blood causes a decrease in  $\text{O}_2$  concentration inside carotid body glomus cells. This causes a decrease in activity of mitochondrial electron transport chain (ETC)<sup>47</sup> and changes in other putative oxygen-sensing pathways, such as oxygen-sensitive  $\text{K}^+$  channels<sup>48,49</sup>, haem oxygenase<sup>50</sup>, AMP kinase<sup>51</sup>, and hydrogen sulphide signalling<sup>37</sup>. These changes are hypothesized to converge on oxygen-sensitive  $\text{K}^+$  channels, which close in hypoxia and depolarize the plasma membrane. Depolarization then opens voltage-gated  $\text{Ca}^{2+}$  channels, leading to an increase in intracellular calcium that stimulates transmitter release to carotid sinus nerve to increase breathing<sup>2</sup>. Mitochondria of carotid body cells are highly sensitive to hypoxia compared with other tissues, as assayed by imaging of mitochondrial membrane potential, NADH levels, and spectral properties<sup>23,52–54</sup>. Drugs and mutations that inhibit the ETC mimic the effect of hypoxia on carotid body activity and breathing<sup>23,55–59</sup>. **e**, Regulation of lactate production by oxygen. In normoxia, pyruvate produced by glycolysis is transported into mitochondria and efficiently used in Krebs cycle to supply electrons to ETC to produce ATP. In hypoxia, lack of oxygen to act as the final electron acceptor limits electron transport, causing pyruvate to build up and become converted to lactate<sup>5,6,27</sup>. The ETC poison cyanide inhibits the haem  $a_3$  subunit of cytochrome  $c$  oxidase to prevent transfer of electrons to oxygen, leading to lactate accumulation even in presence of adequate oxygen<sup>60</sup>. Cytosolic lactate accumulation results in transport of lactic acid (lactate and  $\text{H}^+$ ) out of the cell by monocarboxylate transporters<sup>6,61</sup>. In normoxia, lactate concentrations in blood, tissue, and tissue interstitium are 1–5 mM<sup>62,63</sup>.



**Extended Data Figure 2 | RNA sequencing and whole-genome microarrays detect *Olfr78* transcripts enriched in the carotid body.** **a**, Histogram of frequency of genes for different levels of expression enrichment in carotid body relative to adrenal medulla by RNA sequencing. Values are  $\log_2(\text{carotid body/adrenal medulla})$ , with data binned for every  $\log_2$  interval of 1.0 centred at integers. **b**, Plot of  $\log_2$  values of RPKM in carotid body and adrenal medulla of all 1,126 OR genes annotated in RefSeq shown in alphanumerical order. The five OR genes expressed at RPKM  $> 2$  (dashed line) are indicated. Samples that had no transcripts are plotted at a value of  $-7.1$ , just below the smallest RPKM value for ORs. Data presented in Supplementary Table 1. **c**, Comparison of expression levels of  $> 34,000$  genes in adult mouse carotid body and adrenal medulla by whole-genome microarrays. Plot shows  $\log_2$  of the ratio for carotid body relative to adrenal medulla of the fluorescence intensity values for the 45,000 probe sets. The three probe sets for *Olfr78* transcripts are indicated (circles). Expression of *Olfr78* was significantly different between carotid body and adrenal medulla for all three probe sets

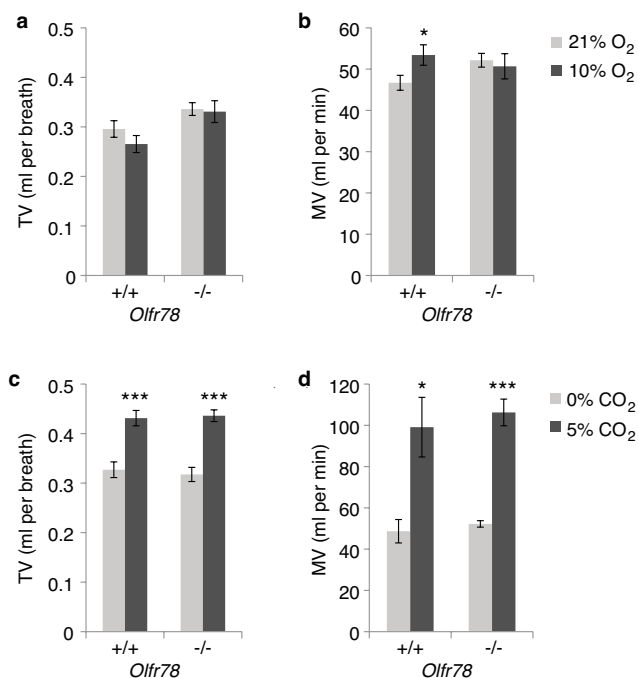
( $P < 0.05$  by ANOVA with control for false discovery rate). **d**, Histogram of the frequency of genes for different levels of expression enrichment in carotid body relative to adrenal medulla in microarray data. Values are  $\log_2(\text{carotid body/adrenal medulla})$ , with data binned for every  $\log_2$  interval of 1.0 centred at integers. The three probe sets detecting *Olfr78* mRNA (arrows) confirmed the RNA sequencing data (**a**, Fig. 1a, b and Extended Data Table 1) showing *Olfr78* among the mRNAs most highly enriched in carotid body. Mouse carotid body *Olfr78* expression is consistent with previous microarray data<sup>7,64</sup>. **a–d**, Three cohorts of ten animals each. Data as mean. **e**, Genomic locus showing the large cluster of  $\sim 160$  class I OR genes on chromosome 7, with region encoding MOR18 subfamily (*Olfr78*, *Olfr558*, and *Olfr557*) expanded below. We did not detect transcripts in either tissue for *Olfr557*, which lies adjacent to *Olfr558* in the cluster, or for the intervening (*Olfr33*, *Olfr559*) and intronic (*Olfr560*) ORs. Clusters of genes encoding globins, Trims, and USP proteins are also found with this OR cluster. Large box, coding sequence; arrowhead, coding orientation; small box, non-coding exons.



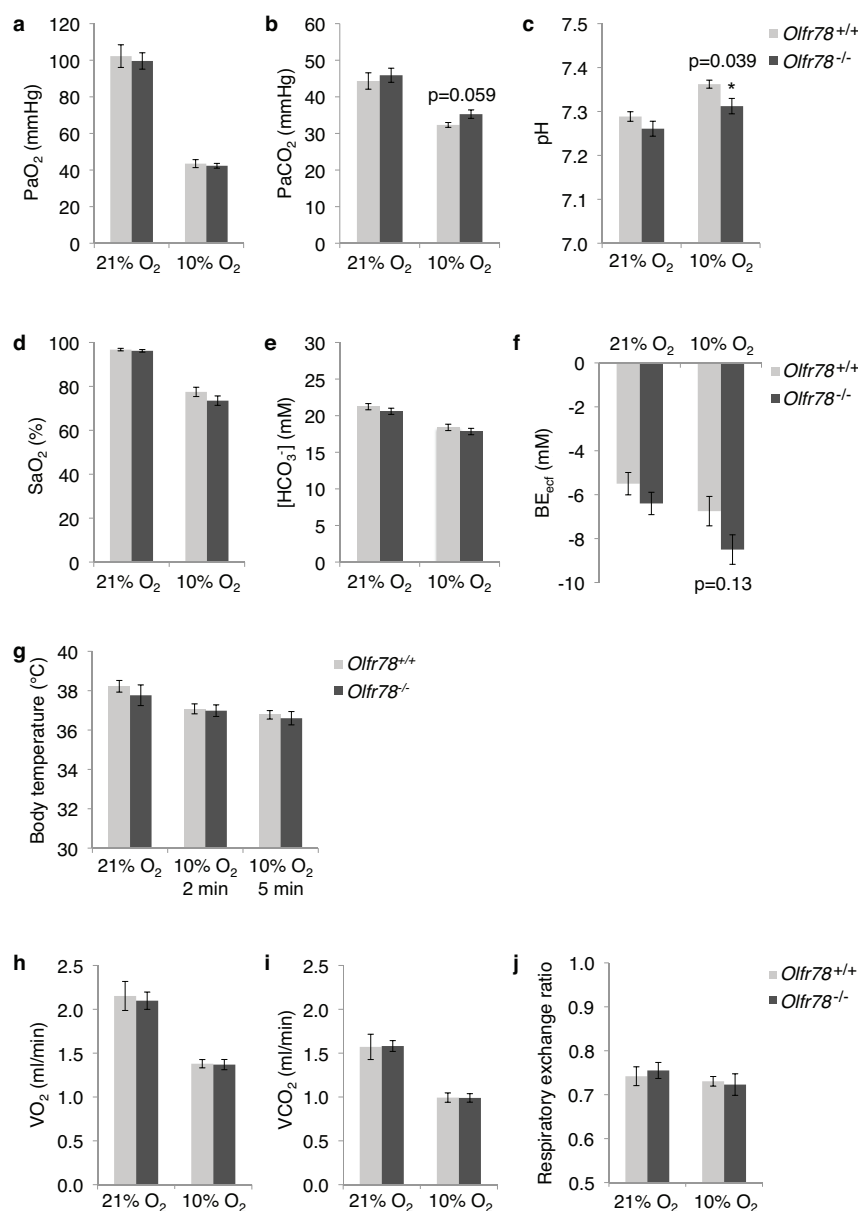


**Extended Data Figure 3 | *Olfr78* and *Olfr558* expression in tissues in the oxygen-sensing circuit.** Expression of *Olfr78* reporter in heterozygous (a) and homozygous (b–d) *Olfr78-GFP-*taulacZ** reporter animals<sup>11</sup>. a–c, Sections of carotid bifurcations stained for GFP (*Olfr78* reporter; green), TH (red), and DAPI (nuclei; blue). a, Section of carotid body showing co-expression of reporter GFP and TH in glomus cells. Monoallelic expression would predict that only half of TH-positive cells express the reporter<sup>12</sup>. Arrowheads, clusters of glomus cells expressing both GFP and TH. b, c, Sections of the same carotid bifurcation. Panels on right show close-ups of boxed region (petrosal ganglion, PG). No GFP-positive cells were found in petrosal ganglion. TH-positive nerve fibres (arrowheads) and cell bodies were found in glossopharyngeal nerve (GN) and petrosal ganglion. Dashed circle indicates vagus nerve

(VN). NG/JG, nodose/jugular ganglia. d, X-gal staining of a brain sagittal section. Reporter expression (blue) was restricted to olfactory bulb (arrowhead) in this section and complete brain serial sagittal sections. Anterior, right; dorsal, up. e–h, *Olfr558* expression in a knockout/reporter mouse in which the *Olfr558* coding region is replaced with *lacZ* encoding β-galactosidase. e, *Olfr558* reporter expression in blood vessels of carotid body and SCG by X-gal staining. Heterozygous *Olfr558*<sup>+/lacZ</sup> samples showed the same pattern of staining (data not shown). f–h, Carotid body sections immunostained for β-galactosidase (*Olfr558* reporter; green) and TH (red) with DAPI counterstain (blue) in f, and for β-galactosidase (green) and CD31 (red) in g or smooth muscle actin (red) in h. Scale bars, 100 μm (a, b right, c right, f–h), 200 μm (b left, c left), 500 μm (e), and 2 mm (d).



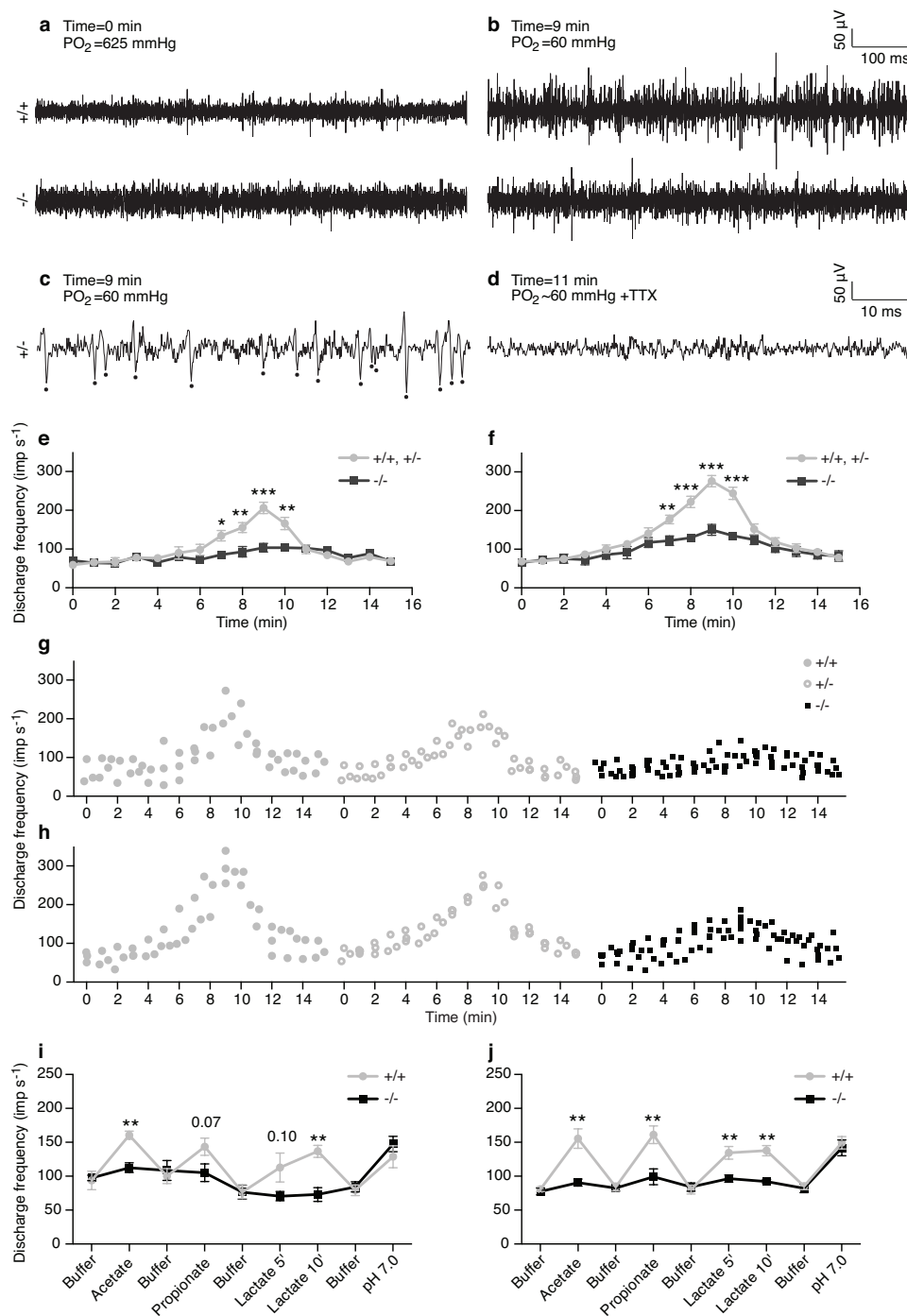
**Extended Data Figure 4 | Tidal volume and minute ventilation of *Olfr78*<sup>-/-</sup> mutants exposed to hypoxia and hypercapnia.** Whole body plethysmography of unrestrained, unanaesthetized *Olfr78*<sup>+/+</sup> control and *Olfr78*<sup>-/-</sup> mutant littermates (as in Fig. 2). **a, b**, Tidal volume (TV) and minute ventilation (MV) of animals exposed to hypoxia. Sample size  $n = 9$  (+/+), 8 (-/-) animals. **c, d**, TV and MV of animals exposed to hypercapnia. Sample size  $n = 4$  (+/+), 5 (-/-) animals. Data as mean  $\pm$  s.e.m. \* $P < 0.05$ , \*\*\* $P < 0.001$  by paired  $t$ -test.



**Extended Data Figure 5 | Physiological responses of  $\text{Olfr78}^{-/-}$  mutants to hypoxia *in vivo*.** **a–f**, Arterial blood gas measurements of  $\text{Olfr78}^{+/+}$  control and  $\text{Olfr78}^{-/-}$  mutant animals exposed to hypoxia. Values are  $\text{PaO}_2$  (**a**),  $\text{PaCO}_2$  (**b**), and pH (**c**) of blood collected from the right carotid artery of anaesthetized  $\text{Olfr78}^{+/+}$  control and  $\text{Olfr78}^{-/-}$  mutant animals exposed to normoxia (21%  $\text{O}_2$ ) and hypoxia (10%  $\text{O}_2$ ) for 3 min. Oxygen saturation ( $\text{SaO}_2$ , **d**),  $[\text{HCO}_3^-]$  (**e**), and base excess of extracellular fluid ( $\text{BE}_{\text{ecf}}$ , **f**) calculated from  $\text{PaO}_2$  (**a**),  $\text{PaCO}_2$  (**b**), and pH (**c**) values. Sample size  $n=4$  ( $+/+$ , 21%  $\text{O}_2$ ), 5 ( $-/-$ , 21%  $\text{O}_2$ ), 4 ( $+/+$ , 10%  $\text{O}_2$ ),

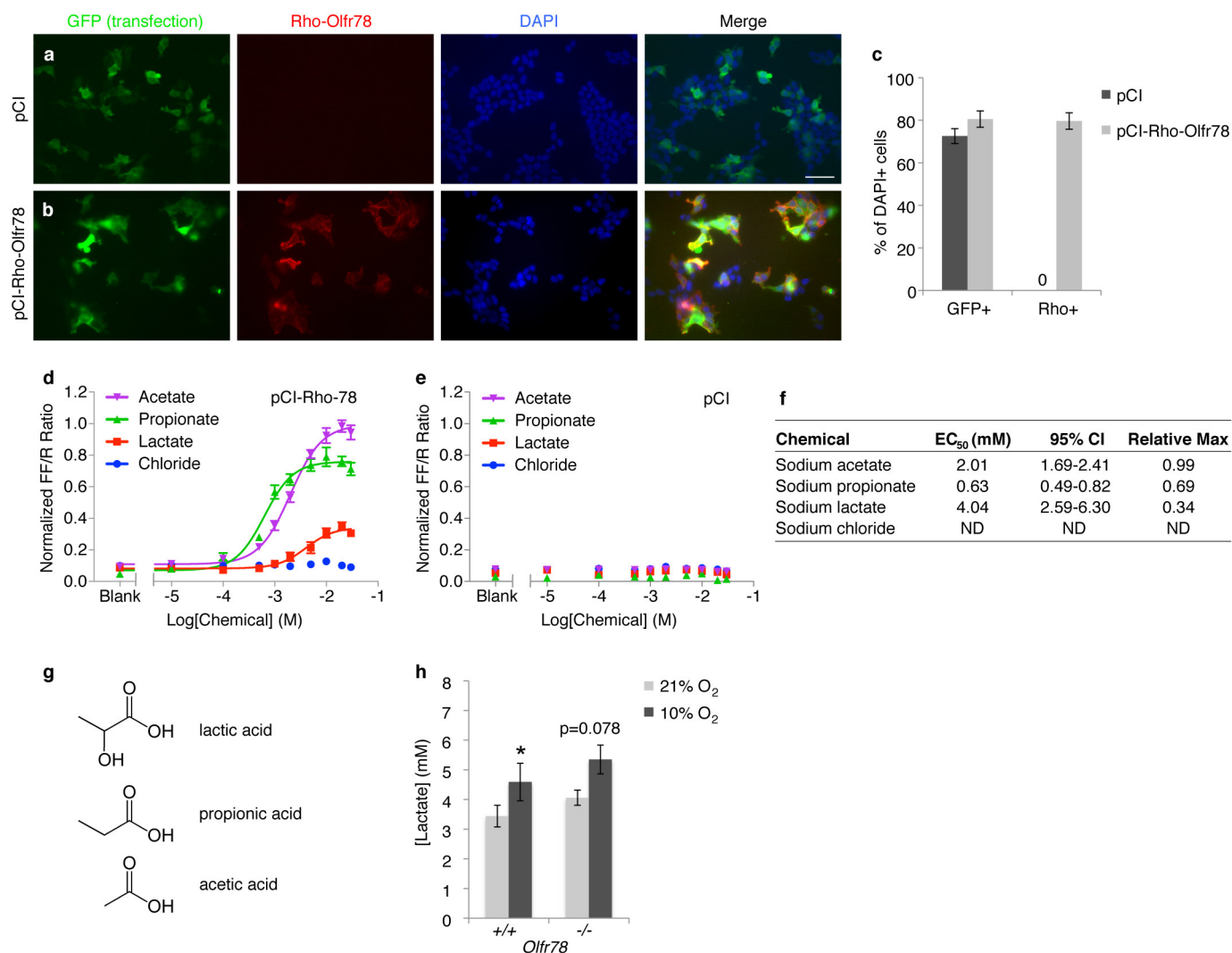
6 ( $-/-$ , 10%  $\text{O}_2$ ) animals. **g**, Body temperature of unanaesthetized  $\text{Olfr78}^{+/+}$  control and  $\text{Olfr78}^{-/-}$  mutant littermates in room air (21%  $\text{O}_2$ ) and exposed to hypoxia (10%  $\text{O}_2$ ) for indicated times. Sample size  $n=4$  ( $+/+$ ), 6 ( $-/-$ ) animals. **h–j**, Metabolic values measured by indirect calorimetry of unanaesthetized  $\text{Olfr78}^{+/+}$  control and  $\text{Olfr78}^{-/-}$  mutant littermates exposed to normoxia (21%  $\text{O}_2$ ) and hypoxia (10%  $\text{O}_2$ ) for 10 min. Sample size  $n=4$  ( $+/+$ ), 6 ( $-/-$ ) animals. Data as mean  $\pm$  s.e.m. \* $P < 0.05$  by unpaired  $t$ -test.





**Extended Data Figure 6 | Carotid body chemosensory responses assayed by carotid sinus nerve activity.** **a, b**, Raw discharge frequency (extracellular recording) of carotid sinus nerves from *Olfr78*<sup>+/+</sup> control and *Olfr78*<sup>-/-</sup> mutant animals at time 0 (**a**) and 9 min (**b**) after the change in gas bubbling the perfusion buffer from 95% O<sub>2</sub>/5% CO<sub>2</sub> to 95% N<sub>2</sub>/5% CO<sub>2</sub>. **c, d**, Carotid sinus nerve activity of an *Olfr78*<sup>+/+</sup> nerve 9 min after the change in gas to 95% N<sub>2</sub>/5% CO<sub>2</sub> (**c**) and 2 min later after addition of 7.5 μM TTX while still bubbling 95% N<sub>2</sub>/5% CO<sub>2</sub> (**d**). Scored action potentials are marked by filled circles. **e-h**, Time course of carotid sinus nerve activity in the *Olfr78* genotypes indicated, scored using Spike2 software (**e, g**) or by hand (**f, h**) and showing mean ± s.e.m. (**e, f**) or individual (**g, h**) values. The residual responses of *Olfr78*<sup>-/-</sup> nerves to hypoxia were more apparent when scored by hand. Sample size *n* = 6 (3 +/+, 3 +/-), 5 (-/-) animals. \**P* < 0.05, \*\**P* < 0.01,

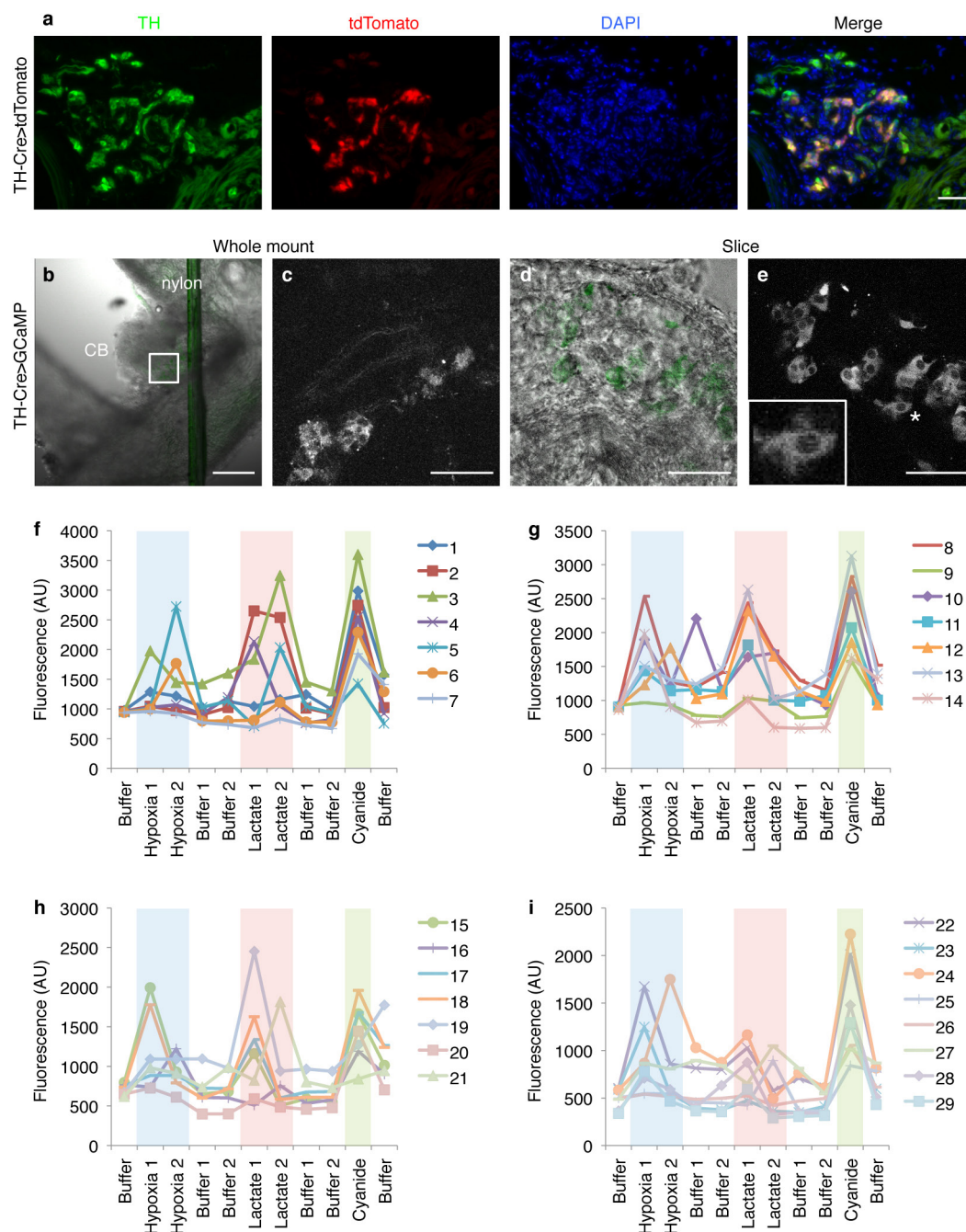
\*\*\**P* < 0.001 by unpaired *t*-test. *Olfr78*<sup>+/+</sup> and *Olfr78*<sup>+/+</sup> recordings were not significantly different from each other at any time point, except for time = 11 min, by unpaired *t*-test (*P* > 0.05). **i, j**, Time course of raw discharge of carotid sinus nerves from *Olfr78*<sup>+/+</sup> control and *Olfr78*<sup>-/-</sup> mutant animals in response to acetate (30 mM, 5 min), propionate (30 mM, 5 min), and lactate (30 mM, 5 and 10 min), and pH 7.0 (5 min) scored using Spike2 software (**i**) or by hand (**j**). Recovery times were 15 min between acetate, propionate, and lactate, and at least 30 min between lactate and pH 7.0. To minimize the contribution of endogenous hypoxic signals, the superfusion buffer in the chamber was maintained at hyperoxic conditions (PO<sub>2</sub> = 625 mmHg). Sample size *n* = 5 (+/+), 5 (-/-) animals. Data as mean ± s.e.m. \**P* < 0.05, \*\**P* < 0.01 by unpaired *t*-test.



**Extended Data Figure 7 | Lactate activates Olfr78 expressed in HEK293T cells and increases acutely in blood in hypoxia *in vivo*.**

**a, b**, HEK293T cells transfected with empty vector pCI (**a**) or pCI-Rho-Olfr78 (**b**) and RTP1S (OR transport protein) and cytoplasmic GFP (co-transfection marker) plasmids. Transfected cells were stained before fixation to detect Rho-tagged Olfr78 (anti-Rho; red) on the cell surface. GFP (transfection marker, green); DAPI (nuclei, blue). Bar, 100  $\mu$ m. **c**, Quantitation of cells expressing GFP and Rho as percentage of DAPI-positive cells in fields shown in **a** and **b**. Sample size  $n = 164$  (pCI), 108 (pCI-Rho-Olfr78) cells. Data as percentage  $\pm$  standard error of percentage. **d**, Dose-response curves for propionate, acetate, and chloride compared with lactate in activation of Olfr78 in transfected HEK293T cells as in Fig. 4a. Sample size  $n = 8$  (propionate), 12 (acetate), 4 (chloride), and 12 (lactate) wells. Data as mean  $\pm$  s.e.m. By analysis of variance

(ANOVA), all chemicals except chloride ( $P = 0.309$ ) showed significant difference ( $P < 0.001$ ). **e**, Dose-response curves as in **c** except cells were transfected with empty vector (pCI). ANOVA showed no significant difference ( $P > 0.05$ ) for any chemical. **f**, EC<sub>50</sub>, 95% confidence interval of EC<sub>50</sub>, and relative maximal activation values from fitted curves in **c**. ND, not determined owing to lack of curve fitting to data. **g**, Structures of the short-chain fatty acids. **h**, Lactate concentrations in blood collected from tail artery of restrained, unanaesthetized *Olfr78*<sup>+/+</sup> control and *Olfr78*<sup>-/-</sup> mutant littermates exposed to hypoxia (10% O<sub>2</sub>) for 4–5 min. Values for animals in normoxia (21% O<sub>2</sub>) are likely to be an overestimate of baseline concentrations owing to greater restraint required to immobilize animals in normoxia<sup>41</sup>. Sample size  $n = 5$  (+/+), 6 (-/-) animals. Data as mean  $\pm$  s.e.m. \* $P < 0.05$  by paired  $t$ -test.



**Extended Data Figure 8 | Calcium imaging of responses of carotid body glomus cells to chemosensory stimuli.** **a**, Carotid body of a *Th-Cre; ROSA-tdTomato* adult immunostained for the Cre-dependent reporter tdTomato (red) and TH (green) to show glomus cells<sup>31,32,34</sup> and counterstained with DAPI (nuclei, blue). tdTomato labelled glomus cells. **b–e**, Tissue preparations for calcium imaging of carotid bodies from *TH-Cre; ROSA-GCaMP3* animals that express the calcium indicator GCaMP3 selectively in glomus cells<sup>31–33</sup>. **b**, Differential interference contrast image of whole mount carotid bifurcation with GCaMP3 fluorescence pseudocoloured green. **c**, High-magnification, two-photon image of boxed region in **b**. **d**, Differential interference contrast image of carotid body tissue slice with GCaMP3 fluorescence pseudocoloured green. **e**, Two-photon image of carotid body slice in **d**. Inset shows glomus cell marked by asterisk at higher magnification. GCaMP3 fluorescence was

seen in cytoplasm and excluded from nucleus of glomus cells. Bars, 100  $\mu$ m (**a**), 200  $\mu$ m (**b**), 50  $\mu$ m (**c–e**). **f–i**, Time course of calcium responses of individual glomus cells to hypoxia, lactate, and cyanide. Whole mount carotid bodies were exposed sequentially to hypoxia (40–50 mmHg), lactate (30 mM), and cyanide (2 mM). Interval between data points is ~2 min, the time required to acquire a stack of images through the carotid body, excluding the 2 min ramp times between stimuli. All glomus cells analysed ( $n = 42$  cells) responded strongly to cyanide. Fluorescence traces shown are the 29 individual glomus cells that responded to both hypoxia and lactate, arranged in order of decreasing initial fluorescence intensity. The other 13 glomus cells responded either to hypoxia (9 cells) or to lactate (4 cells). Multiple data points for buffer or stimuli were averaged to generate the data presented in Fig. 4c. Background colours match bar colours in Fig. 4d.

Extended Data Table 1 | Top 150 genes highly expressed in carotid body versus adrenal medulla by RNA sequencing

Gene Name	CB*	AM*	CB/AM†	Comment	Gene Name	CB*	AM*	CB/AM†	Comment
Cpne4	9.9	0.4	9.4	Ca <sup>2+</sup> -dep membrane binding	Cpa3	9.4	3.7	5.7	Peptidase, mast cell
Slc9a2	9.2	0.3	8.9	Na <sup>+</sup> /K <sup>+</sup> exchanger	Dlx6	5.7	0.0	5.7	Hox TF, brain
Coch	10.1	1.2	8.9	Expressed in ear cells	2210407C18Rik	5.2	-0.6	5.7	
Abcc2	9.0	0.6	8.4	ABC transporter	Cwh43	7.0	1.3	5.7	GPI-anchor lipid remodeling
Barx1	7.7	-0.6	8.3	Hox TF, neural crest	Cma1	9.8	4.2	5.7	Chymase, mast cell
Higd1c	11.3	3.2	8.1	Membrane protein	Nmu	5.0	-0.6	5.6	Neuropeptide, neuromedin U
Hmx3	7.3	-0.6	7.8	Hox TF, ear and neurons	Drd1a	5.8	0.3	5.5	GPCR, dopamine receptor 1a
Hmx2	7.6	-0.2	7.8	Hox TF, ear	Sfrp4	9.8	4.3	5.5	Secreted frizzled
Ccl21a	8.2	0.4	7.8	Chemokine	Bcl11b	8.6	3.1	5.5	Tumor suppressor
AB099516	13.4	5.6	7.8	Methyltransferase-like	Myh11	10.6	5.1	5.5	Smooth muscle myosin
Pcolce2	11.3	3.6	7.6	Procollagen peptidase enhancer	Ptx4	4.9	-0.6	5.5	Multifunctional conserved
Lypd2	8.3	0.7	7.5	LY6/PLAUR domain containing	Rxrg	8.7	3.3	5.5	Retinoid X receptor, nuclear
Gkn3	8.7	1.3	7.3	Pseudogene	Pax8	4.9	-0.6	5.5	Paired box TF, thyroid
Mcpt4	8.8	1.4	7.3	Mast cell proliferation	Chrhr1	5.7	0.3	5.4	GPCR, hormone
Gm21541	7.3	0.0	7.3	Chemokine	Gm694	4.9	-0.6	5.4	
Ccl21b	7.3	0.0	7.3	Chemokine	Mc4r	7.5	2.1	5.4	GPCR, melanocortin receptor
Gm13304	7.4	0.1	7.3	Chemokine	Sgca	5.3	0.0	5.4	Extracellular matrix
Tpsb2	7.8	0.6	7.1	Tryptase, mast cell	Gm9885	5.4	0.1	5.3	
Cpne5	8.1	1.0	7.1	Ca <sup>2+</sup> -dep membrane binding	Kcna1	12.4	7.1	5.3	Voltage-gated K <sup>+</sup> channel, K <sub>v</sub> 1.1
H19	8.7	1.7	7.1	Maternal imprinted	Smim5	8.0	2.7	5.3	Integral membrane protein
Gm10768	7.1	0.1	7.0		Ccl11	9.0	3.8	5.3	Chemokine
Gm1987	8.2	1.3	6.9	Chemokine	Plip	6.6	1.3	5.2	Involved in myelination
LOC100041504	7.2	0.3	6.9	Chemokine	Galnt9	8.8	3.6	5.2	GalNac transferase
Cst12	6.6	-0.2	6.9	Cysteine endopeptidase	C1qtnf9	6.9	1.7	5.2	Adipokine
Vit	9.3	2.4	6.9	Extracellular matrix	Hcn1	6.6	1.4	5.2	Hyperpol, CNG K <sup>+</sup> channel
Mpz	9.0	2.1	6.8	Myelin	Prss12	6.8	1.6	5.2	Trypsin, neuronal plasticity
Ccl21c	7.2	0.4	6.8	Chemokine	Rergl	6.1	0.9	5.2	Putative GTPase
Prdm6	6.3	-0.6	6.8	Histone methyltransferase	4931429I11Rik	6.3	1.1	5.2	
Susd5	8.1	1.3	6.8	Sushi domain containing	Enpp3	6.8	1.6	5.2	Nucleotide ectoenzyme
Rxfp1	6.6	-0.2	6.8	GPCR, relaxin receptor	Saa1	5.3	0.1	5.2	Serum amyloid
Cytl1	10.1	3.3	6.8	Cytokine-like	Itgb4	9.3	4.2	5.1	Integrin
Olfir558	8.5	1.8	6.8	GPCR, olfactory receptor	Cpn1	4.5	-0.6	5.1	Carboxypeptidase
Ly6h	11.1	4.4	6.8	Lymphocyte antigen complex	Cspg4	9.6	4.6	5.0	Extracellular matrix proteoglycan
Gpr139	6.2	-0.6	6.7	GPCR	9130206I24Rik	7.2	2.2	5.0	
C530044C16Rik	7.7	1.0	6.7		Areg	4.4	-0.6	5.0	EGF/TGF- $\alpha$ homolog, glia
Olfir78	11.7	5.1	6.7	GPCR, olfactory receptor	Trabd2b	8.9	3.9	5.0	Neg regulator of Wnt signaling
Sfrp2	10.3	3.6	6.6	Secreted frizzled-related	Chrdl1	9.1	4.2	5.0	BMP4 antagonist, neuron
Pln	9.4	2.7	6.6	Ca <sup>2+</sup> -ATPase inhibitor, heart	Ltbp2	7.3	2.4	5.0	Extracellular matrix
Tph1	7.4	0.7	6.6	Serotonin biosynthesis	Rbp7	7.0	2.1	4.9	Retinol binding protein
Gm10591	7.3	0.6	6.6	Chemokine	Mmrn1	7.9	3.0	4.9	Carrier of platelet proteins
Cntnap4	6.0	-0.6	6.5	Neural cell adhesion	Hrk	8.0	3.1	4.9	Promotes apoptosis
LOC100041593	7.3	0.8	6.5		Ptgr	6.8	1.9	4.9	GPCR, prostaglandin receptor
Thbs4	7.3	0.8	6.5	Cell adhesion glycoprotein	Tpsab1	5.0	0.1	4.9	Tryptase
Tnmd	6.3	-0.2	6.5	Cartilage-specific glycoprotein	Pthlh	7.7	2.9	4.9	Parathyroid hormone-like
Gpr20	7.1	0.6	6.5	GPCR	Fndc1	8.9	4.0	4.9	Fibronectin domain containing
Cntn5	5.9	-0.6	6.5	Neural cell adhesion	Bmx	6.6	1.8	4.8	Non-receptor tyrosine kinase
Dgkh	11.2	4.8	6.4	Diacylglycerol kinase	Car12	8.7	3.9	4.8	Carbonic anhydrase, extracellular
Xirp1	5.9	-0.6	6.4	Actin binding	Atp8b1	7.0	2.2	4.8	Cation transport ATPase
Ptx3	7.1	0.8	6.4	Innate immunity, inflammation	Cd1631	4.6	-0.2	4.8	Scavenger receptor, immune
Prom1	9.1	2.7	6.4	Adult stem cell maintenance	Gcsam	5.4	0.6	4.8	PDZ domain containing
Nov	9.2	2.8	6.4	Extracellular matrix, cancer	Acta2	13.9	9.1	4.8	Smooth muscle actin, aorta
Adamts13	6.9	0.6	6.3	Metalloprotease, cancer	Dcn	15.1	10.3	4.8	Extracellular matrix proteoglycan
Gdnf	7.1	0.7	6.3	Glial derived neurotrophic factor	Dlx6as1	7.2	2.4	4.8	Noncoding RNA, GABA neurons
Dlx5	7.5	1.3	6.3	Hox TF, bone	Chp2	6.9	2.2	4.7	Ca <sup>2+</sup> -binding, pH control
Inmt	11.8	5.7	6.2	Indolethylamine N-methylase	Gm10808	5.2	0.4	4.7	
Mfap5	11.5	5.4	6.1	Microfibril-associated glycoprotein	Col8a1	10.1	5.4	4.7	Collagen
Scara3	9.4	3.3	6.1	ROS scavenger	Foxd1	6.0	1.3	4.7	Forkhead TF
Pgf	10.0	3.9	6.0	Placental growth factor	Igfbp4	13.2	8.5	4.7	Insulin-like growth factor binding
Mustn1	10.0	4.0	6.0	Expressed in muscle and bone	Alpl	8.5	3.8	4.7	Alkaline phosphatase
Pkib	12.5	6.5	6.0	cAMP-dep kinase inhibitor	Dhrs2	6.2	1.5	4.7	Dehydrogenase/reductase
Comp	6.6	0.6	6.0	Cartilage matrix	Lmod3	4.5	-0.2	4.7	Expressed in muscle
Gap43	12.3	6.3	6.0	Neuronal growth cone	Rnase1	5.1	0.4	4.7	Secretory RNase
Shisa3	8.7	2.7	6.0	Fgf and Wnt signaling interactor	Gm15998	4.4	-0.2	4.7	
Cma2	5.4	-0.6	6.0	Chymase, mast cell	Cnn1	7.4	2.7	4.7	Thin filament associated, muscle
Gjb5	6.1	0.1	6.0	Gap junction	Col2a1	4.1	-0.6	4.7	Collagen
Meox2	9.1	3.2	5.9	Hox TF, mesenchyme	Osr1	8.2	3.6	4.7	TF, odd-skipped related
Plch1	5.7	-0.2	5.9	Phospholipase	Hepacam2	6.0	1.3	4.6	Centrosome maturation
Tlx1	5.3	-0.6	5.9	Hox TF-spleen, neurons, T cells	Fcer1a	5.1	0.5	4.6	IgE receptor, Fc
Igfbp6	12.1	6.3	5.8	Insulin-binding protein	Pax9	4.1	-0.6	4.6	Paired box TF
Rgs5	14.9	9.0	5.8	GTPase activator	Ncmap	7.5	2.8	4.6	Myelin-associated
Edn1	8.0	2.2	5.8	Endothelin 1, vasoconstrictor	Pls1	5.5	0.8	4.6	Plastins, hematopoietic cells
Rspo1	10.1	4.2	5.8	Activator of Wnt signaling	Fbln7	11.1	6.5	4.6	Extracellular matrix adhesion
Upb1	7.0	1.2	5.8	Pyrimidine degradation	Folr2	8.4	3.8	4.6	Folate receptor
Scnn1b	9.4	3.6	5.8	ENaC Na <sup>+</sup> channel, beta subunit	Mxs1	6.9	2.3	4.6	Hox, craniofacial
Fgf7	6.6	0.8	5.8	Fibroblast growth factor	Tmem158	12.7	8.1	4.6	Peptide receptor

\*Carotid body and adrenal medulla values are log<sub>2</sub>(aligned reads per 10<sup>7</sup> reads) as Fig. 1a. †Values are log<sub>2</sub>(carotid body/adrenal medulla) ratios as in Extended Data Fig. 2a. All genes were significantly different between carotid body and adrenal medulla by paired *t*-test ( $P < 0.05$ ), except *Bmx* ( $P = 0.056$ ). Yellow highlight, genes for ORs. Blue highlight, genes previously shown to be expressed in carotid body (refs 40, 65, 66). TF, transcription factor; GPCR, G-protein-coupled receptor; ROS, reactive oxygen species; CNG, cyclic nucleotide-gated.



Extended Data Table 2 | Expression of genes associated with olfactory neurons

Gene Name	Function	RPKM		
		CB*	AM*	CB/AM
<i>Olf78</i>	olfactory receptor	101	1.1	92 <sup>†</sup>
<i>Olf558</i>	olfactory receptor	13	0.13	102 <sup>‡</sup>
<i>Gnal</i>	G protein, alpha subunit	11	3.9	2.8 <sup>†</sup>
<i>Adcy3</i>	adenylate cyclase	3.2	5.2	0.61 <sup>†</sup>
<i>Cnga2</i>	cyclic nucleotide-gated channel	0.060	0.10	0.61
<i>Cnga4</i>	cyclic nucleotide-gated channel	0.075	0.053	1.4
<i>Cngb1</i>	cyclic nucleotide-gated channel	0.23	0.23	1.0
<i>Ano2</i>	calcium-activated chloride channel	0.75	0.11	6.8 <sup>†</sup>
<i>Ric8b</i>	guanine nucleotide exchange factor	8.6	8.8	1.0
<i>Rtp1</i>	olfactory receptor trafficking	0.082	0.11	0.77
<i>Rtp2</i>	olfactory receptor trafficking	0.048	0.047	1.0
<i>Reep1</i>	olfactory receptor trafficking	6.2	4.3	1.4
<i>Omp</i>	mature olfactory neuron marker	4.0	3.5	1.1

\*Carotid body and adrenal medulla values are RPKM normalized to longest mRNA isoform for each gene.

<sup>†</sup> $P < 0.05$  between carotid body and adrenal medulla by paired *t*-test.

<sup>‡</sup> $P < 0.01$  between carotid body and adrenal medulla by paired *t*-test.

# Resensitizing daclatasvir-resistant hepatitis C variants by allosteric modulation of NS5A

Jin-Hua Sun<sup>1</sup>, Donald R. O'Boyle II<sup>1</sup>, Robert A. Fridell<sup>1</sup>, David R. Langley<sup>2</sup>, Chunfu Wang<sup>1</sup>, Susan B. Roberts<sup>1</sup>, Peter Nower<sup>1</sup>, Benjamin M. Johnson<sup>3</sup>, Frederic Moulin<sup>3</sup>, Michelle J. Nophsker<sup>3</sup>, Ying-Kai Wang<sup>4</sup>, Mengping Liu<sup>1</sup>, Karen Rigat<sup>1</sup>, Yong Tu<sup>5</sup>, Piyasena Hewawasam<sup>5</sup>, John Kadow<sup>5</sup>, Nicholas A. Meanwell<sup>5</sup>, Mark Cockett<sup>1</sup>, Julie A. Lemm<sup>1</sup>, Melissa Kramer<sup>3</sup>, Makonen Belema<sup>5</sup> & Min Gao<sup>1</sup>

**It is estimated that more than 170 million people are infected with hepatitis C virus (HCV) worldwide<sup>1,2</sup>. Clinical trials have demonstrated that, for the first time in human history, the potential exists to eradicate a chronic viral disease using combination therapies that contain only direct-acting antiviral agents<sup>3</sup>. HCV non-structural protein 5A (NS5A) is a multifunctional protein required for several stages of the virus replication cycle<sup>4</sup>. NS5A replication complex inhibitors, exemplified by daclatasvir (DCV; also known as BMS-790052 and Daklinza), belong to the most potent class of direct-acting anti-HCV agents described so far, with *in vitro* activity in the picomolar (pM) to low nanomolar (nM) range. The potency observed *in vitro* has translated into clinical efficacy, with HCV RNA declining by ~3–4 log<sub>10</sub> in infected patients after administration of single oral doses of DCV. Understanding the exceptional potency of DCV was a key objective of this study. Here we show that although DCV and an NS5A inhibitor analogue (Syn-395) are inactive against certain NS5A resistance variants, combinations of the pair enhance DCV potency by >1,000-fold, restoring activity to the pM range. This synergistic effect was validated *in vivo* using an HCV-infected chimaeric mouse model. The cooperative interaction of a pair of compounds suggests that NS5A protein molecules communicate with each other: one inhibitor binds to resistant NS5A, causing a conformational change that is transmitted to adjacent NS5As, resensitizing resistant NS5A so that the second inhibitor can act to restore inhibition. This unprecedented synergistic anti-HCV activity also enhances the resistance barrier of DCV, providing additional options for HCV combination therapy and new insight into the role of NS5A in the HCV replication cycle.**

HCV NS5A interacts with viral and cellular proteins, spurring investigations into its role(s) in the virus life cycle<sup>5</sup>. NS5A crystallizes as a dimer in two forms (reviewed in ref. 6), hypothesized to assemble into helical polymers with alternating crystallographic interfaces<sup>7–9</sup>. The pM potency of the NS5A inhibitor DCV (Extended Data Fig. 1), which is unprecedented for antiviral agents<sup>10</sup>, triggered an investigation of the mechanism of DCV inhibition.

We estimated the amount of NS5A protein to be approximately 10.6 fg per cell or ~189 nM (Extended Data Fig. 2). This amount of NS5A is significantly lower than other reported values<sup>11,12</sup>, probably owing to differences in the status of replicon cells maintained in different laboratories. On the basis of our calculations, the ratio of NS5A to DCV in cells is approximately 47,000:1 (Extended Data Fig. 2). This ratio suggests that a small number of inhibitor molecules can impact the function of a large number of NS5A protein molecules, consistent with published observations that NS5A can dimerize in cells, and may form oligomers<sup>8,13</sup>. On the basis of these findings we developed a working model for NS5A inhibitor action: NS5A proteins interact with each

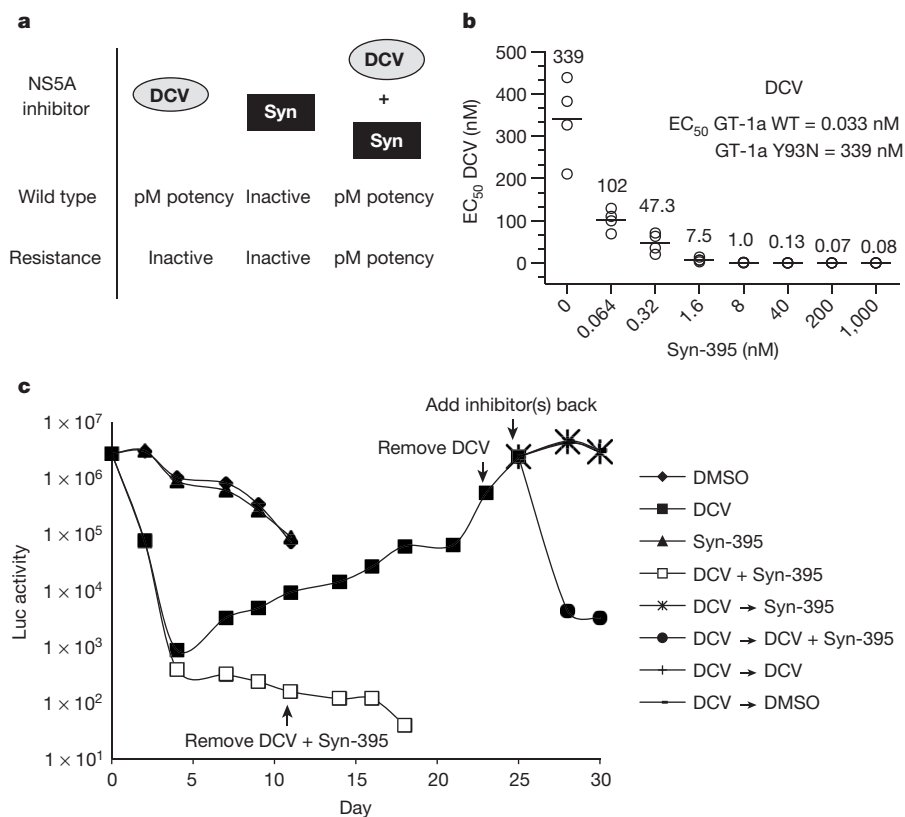
other, and a single bound inhibitor perturbs the function of an NS5A oligomer, disrupting the formation of the replication complex or the function of NS5A within the replication complex and amplifying the inhibitory effect. A methodology combining chemical and classical genetics was used to test this working model.

A biotinylated compound, BMS-671 (Extended Data Fig. 1)<sup>10</sup>, inhibited GT-1b wild-type and resistant (Y93H) replicons with the dramatically different half-maximum effective concentration (EC<sub>50</sub>) values of 33 nM and ~7,400 nM, respectively, an ~224-fold difference (Extended Data Fig. 3a). However, the ability of BMS-671 to bind wild-type and resistant NS5A was very similar (compare BMS-671-bound wild-type and resistant (Y93H) NS5A; Extended Data Fig. 3a)<sup>13</sup>. Moreover, the binding of BMS-671 was competed by the NS5A inhibitor BMS-665 (Extended Data Fig. 1), but only an ~3-fold difference in binding affinity was observed between the wild-type and Y93H variant (3.9 versus 1.4 μM or 7.8 versus 2.8 μM; Extended Data Fig. 3b, c). The discrepancy between potency and binding affinity prompted us to probe the underlying mechanism. We speculated that an NS5A inhibitor such as DCV might bind to resistant NS5A without triggering inhibition, and that a second inhibitor added to the system could deliver one of two functional outcomes: (1) compete with DCV for binding and have no impact on potency; or (2) bind to an adjacent or nearby NS5A that has undergone a conformational change promoted by DCV binding, and enhance potency. Pairs of inhibitors were screened to determine whether either competition or enhancement of activity could be detected.

A combination experiment with pairs of NS5A inhibitors is illustrated in Fig. 1a. In light of the potent EC<sub>50</sub> of DCV on GT-1a wild-type replicons (0.033 nM), EC<sub>50</sub> values >100 nM were considered inactive in this study<sup>14</sup>. The second inhibitor, referred to as a synergist (Syn), is inactive alone against both wild-type and resistant variants but can greatly enhance the potency of DCV against resistant variants. Specifically, DCV exhibits an EC<sub>50</sub> of 0.033 nM against wild-type GT-1a, but has no activity towards a GT-1a Y93N mutant (EC<sub>50</sub> 339 nM; Fig. 1b). The synergist Syn-395 (Extended Data Fig. 1) is inactive towards both wild type and Y93N (EC<sub>50</sub> 214 nM and 215 nM, respectively; Extended Data Fig. 4). However, in the presence of Syn-395, the potency of DCV against the Y93N variant is greatly enhanced. For example, no inhibition of Y93N was observed at 40 nM Syn-395, but in the presence of 40 nM Syn-395 the potency of DCV against Y93N is enhanced by approximately 2,600-fold, with the EC<sub>50</sub> value shifting from 339 nM to 0.13 nM (Fig. 1b). A similar synergistic effect was observed in a reciprocal experiment (Extended Data Fig. 4).

The synergistic effects on potency encouraged us to test the impact of the combination of DCV and Syn-395 on the DCV resistance barrier

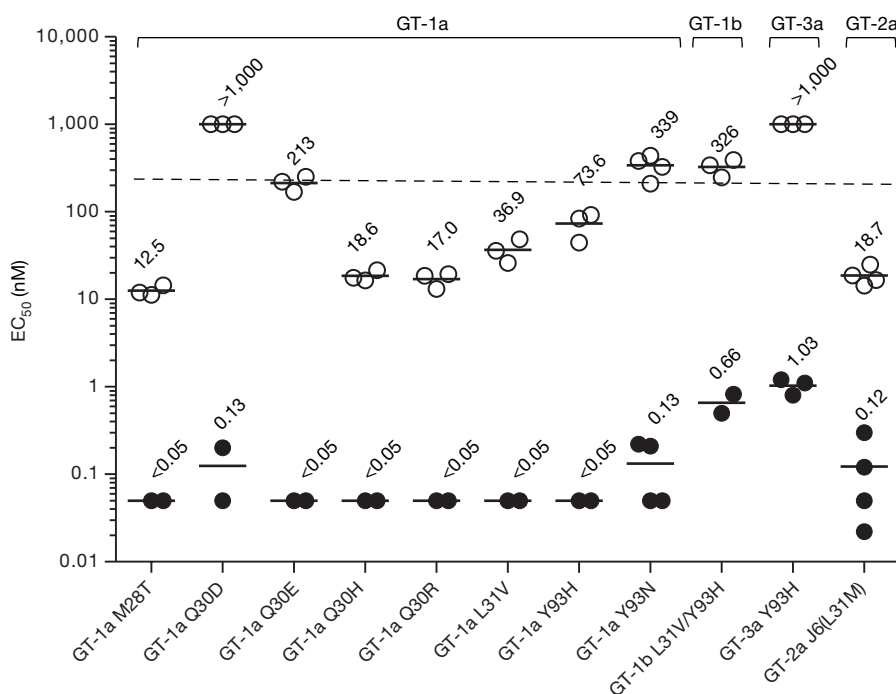
<sup>1</sup>Department of Virology, Bristol-Myers Squibb Research and Development, 5 Research Parkway, Wallingford, Connecticut 06492, USA. <sup>2</sup>Computer-Assisted Drug Design, Bristol-Myers Squibb Research and Development, 5 Research Parkway, Wallingford, Connecticut 06492, USA. <sup>3</sup>Pharmaceutical Candidate Optimization, Bristol-Myers Squibb Research and Development, 5 Research Parkway, Wallingford, Connecticut 06492, USA. <sup>4</sup>Leads Discovery and Optimization, Bristol-Myers Squibb Research and Development, 5 Research Parkway, Wallingford, Connecticut 06492, USA. <sup>5</sup>Discovery Chemistry, Bristol-Myers Squibb Research and Development, 5 Research Parkway, Wallingford, Connecticut 06492, USA.



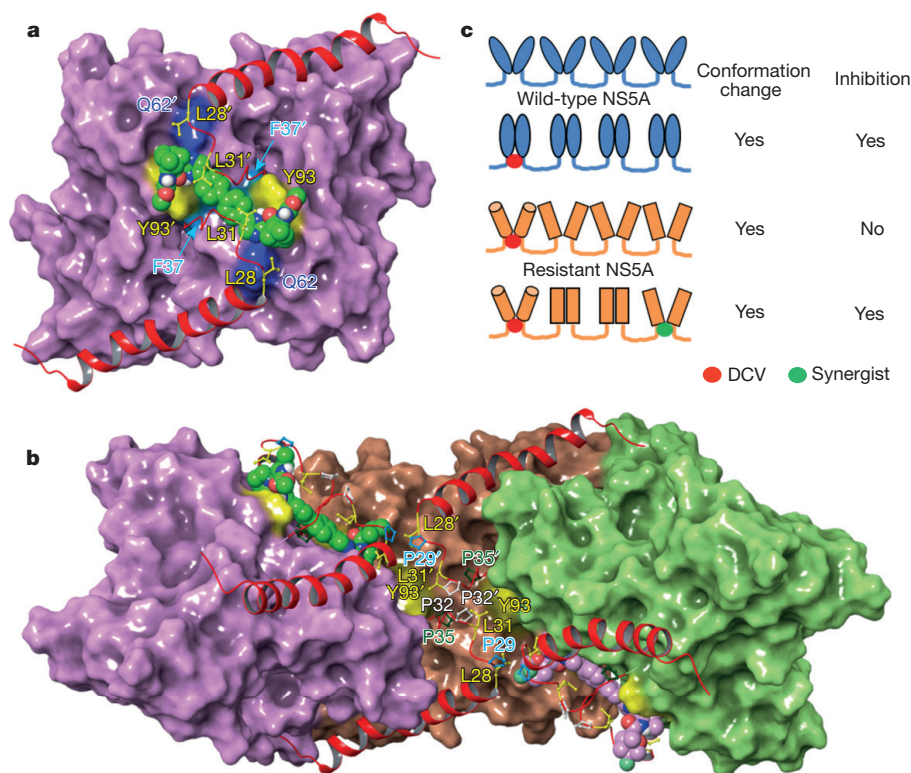
**Figure 1 | Synergistic anti-HCV effects of NS5A compounds.** **a**, Cartoon representation of the synergistic effect of a pair of NS5A compounds. **b**, EC<sub>50</sub> values of DCV determined in the absence or presence of Syn-395 (titrated). Individual values and the average from four independent experiments are shown. WT, wild type. **c**, Huh7 cells infected with GT-1a hybrid virus in the presence of DMSO, DCV (250 nM), Syn-395 (150 nM) or DCV plus Syn-395 combination (250 + 150 nM) were monitored for 30 days. At day 12, DCV and Syn-395 were removed from cells treated with the combination. At day 23, DCV was removed from cells receiving DCV monotherapy; a portion of cells were used for genotypic analysis; the remaining cells were maintained in inhibitor-free medium for 2 days. On day 25, DMSO, DCV (250 nM), Syn-395 (150 nM) and DCV plus Syn-395 (250 + 150 nM) were added to cells that survived DCV monotherapy. Data are representative of two independent experiments.

using an *in vitro* colony elimination assay (Extended Data Fig. 5a)<sup>15</sup>. Consistent with its EC<sub>50</sub>, DCV (30 nM) reduced wild-type GT-1a colony formation dramatically but 100 nM Syn-395 had little or no effect. However, a combination of 10 nM DCV and 100 nM Syn-395 completely eliminated replicons. Genotypic and phenotypic analysis revealed linked substitutions (M28T–Q30R, M28T–S38F or M28T–Q30R–S38F) with higher level resistance in 100% of the replicon cells surviving treatment with the DCV plus Syn-395 combination, indicating that the combination has an enhanced resistance barrier compared to DCV alone.

To investigate the synergistic effect further, a more biologically relevant system—an infectious virus assay—was used to monitor viral kinetics (Fig. 1c). In the absence of compounds (dimethylsulfoxide (DMSO) control), infectious GT-1a HCV replicated well, with a gradual decrease of luciferase (Luc) signal due to the death of infected cells. Similarly, Syn-395 (150 nM) has no apparent inhibitory activity towards wild-type virus. In contrast, infected cells treated with 250 nM DCV (mean trough plasma concentration recorded in phase 1 studies after a single daily dose of 60 mg) displayed viral kinetics similar to those observed in clinical studies with DCV monotherapy:



**Figure 2 | Synergistic effect of DCV and Syn-395 on multiple resistance variants and genotypes.** EC<sub>50</sub> values of DCV were determined in the absence (open circles) and presence (filled circles) of 40 nM Syn-395, using resistant replicon cells of different genotypes. Individual values and the average from  $\geq 2$  independent experiments are shown. The dashed line (250 nM) represents the mean trough plasma concentration of DCV in phase 1 clinical studies (60 mg once daily).



**Figure 3 | Model of DCV and Syn-395 bound to the NS5A dimer and polymers.** **a**, DCV is docked across the NS5A dimer interface such that its hydrophobic biphenyl core sits symmetrically over F37 and F37' (sky blue), and its imidazole rings lie adjacent to Q62 and Q62' (royal blue). The Protein Data Base (PDB) accession 1ZH1 dimer is shown as a mauve surface. N termini amphipathic  $\alpha$ -helices (PDB accession 1R7C) are red; sites of resistance are yellow; carbon atoms of DCV are green. **b**, The NS5A helical hexamer is composed of three PDB accession 1ZH1 dimer interfaces (same colour for each) and two PDB accession 3FQM dimer interfaces (colours alternate). Carbons of DCV and Syn-395 are green

a rapid initial viral decline ( $>3 \log_{10}$ ) (ref. 10), indicated by a decreasing Luc signal between days 0 and 3, followed by an increasing Luc signal due to the emergence of resistance in the presence of DCV (viral breakthrough). Genotypic analysis of the breakthrough virus population showed the emergence of variants (Y93N and Y93R) conferring a relatively high level of resistance ( $>10,000$ -fold) to DCV (ref. 14 and R.A.F., unpublished observations). However, the combination of 250 nM DCV and 150 nM Syn-395 completely suppressed HCV replication. Both compounds were removed at day 12. No viral rebound was observed at days 14, 16 or 18, indicating that HCV was eradicated (Fig. 1c).

The effectiveness of the DCV and Syn-395 cooperative interaction was further investigated by removing DCV on day 23 from the DCV-only treated cultures, when breakthrough (resistant) virus titre was close to baseline level. At day 25, DCV alone, Syn-395 alone or DCV plus Syn-395 were added back to the infected cells (Fig. 1c). Even though DCV and Syn-395 monotherapy yielded no detectable anti-HCV activity, DCV plus Syn-395 yielded an HCV decline of  $\sim 3 \log_{10}$ , with kinetics similar to the initial viral decline. This is a clear illustration of the cooperative interaction between DCV and Syn-395 on DCV-resistant NS5A, resulting in the inhibition of HCV replication.

Another pair of compounds, BMS-393 (ref. 16) and Syn-776 (Extended Data Fig. 1), with acceptable pharmacokinetic profiles in mice was used to test for synergy *in vivo*. The *in vitro* virology profile of this pair is shown in Extended Data Fig. 5b (see also ref. 16). PXB mice<sup>17</sup> were treated with BMS-393 and/or Syn-776 in combination or singly at levels determined to be tolerable (Extended Data Fig. 6a, b). We observed a pattern of activity with a recognizable phenotype: BMS-393

and mauve, respectively. Conformational perturbations induced by DCV into the wild-type NS5A polymer propagate along the helical axis via P29–P35 loop interactions to inhibit multiple NS5A proteins. Smaller (non-inhibitory) conformational changes caused by DCV or Syn-395 individually in resistant NS5A are amplified into potent inhibitory changes when acting together. **c**, DCV binds to resistant NS5A (conformation is different from wild type), without disrupting NS5A function(s); however, DCV binding causes a conformational change that accommodates the binding of a second inhibitor (Syn) on adjacent NS5A molecules to disrupt function(s) of the oligomer.

plus Syn-776 ( $0.4 \text{ mg kg}^{-1}$  plus  $15 \text{ mg kg}^{-1}$ ) yielded a delay in viral rebound compared to either compound alone (Extended Data Fig. 6c). A higher barrier to resistance was supported by the genotypic analysis of variants that emerged during treatment (Extended Data Fig. 6c and Extended Data Table 1).

To determine whether a synergistic combination (DCV and Syn-395) has enhanced potency against additional genotypes (beyond GT-1) and resistant variants, a panel of GT-1a, -1b, -2a and -3a resistant variants that have been observed frequently *in vitro* and in clinical studies<sup>10,18,19</sup> was investigated (Fig. 2). DCV  $\text{EC}_{50}$  values against this panel of GT-1a resistant variants range from 12.5 to  $>1,000 \text{ nM}$ ; the presence of 40 nM Syn-395 enhanced the values to  $<0.2 \text{ nM}$ . Similarly, DCV  $\text{EC}_{50}$  values against GT-1b, GT-2a (J6) and GT-3a resistant variants were enhanced from 326, 18.7 and  $>1,000 \text{ nM}$  to 0.66, 0.12 and 1.03 nM, respectively (Fig. 2). This finding suggests that the feature(s) of NS5A that promotes the synergistic effects characterized extensively for GT-1a is conserved across genotypes. In addition, when Syn-395 was tested in combination with a different NS5A inhibitor, ledipasvir (LDV; also known as GS-5885), a synergistic effect was observed, except for GT-1b L31V–Y93H (Extended Data Fig. 7).

To determine whether a synergistic combination of NS5A inhibitors can be as effective as a combination of direct-acting antiviral agents (DAAs) targeting different HCV proteins, a head-to-head comparison was performed with individual DAAs that have achieved a high sustained viral response as triple combination therapy in phase II clinical studies: DCV 3DAA contains DCV plus an NS3 protease inhibitor (asunaprevir (ASV)) plus an NS5B polymerase inhibitor (beclabuvir (BCV))<sup>20</sup>. Combinations that replace either ASV or BCV with Syn-395



were compared to the DCV 3DAA combination in GT-1a replicon elimination studies. Each of the triple combinations was effective at eliminating the HCV replicon (Extended Data Fig. 8), suggesting that an NS5A synergist could replace either the protease or the polymerase inhibitor in a triple combination. The broad genotype coverage observed with NS5A synergy combinations enhances the potential value of this approach for new HCV therapies.

Several models of NS5A inhibitor–protein complexes can explain the synergistic interactions between pairs of NS5A inhibitors. The model illustrated in Fig. 3 is supported crystallographically. Domain I of NS5A has been crystallized in different forms<sup>7–9</sup> and the structure of the amino-terminal amphipathic  $\alpha$ -helix has been solved by NMR<sup>21</sup>. The structure of the NS5A domain I monomer found in each crystallographic form is highly conserved but presents a unique symmetrical dimer interface. These forms are hypothesized to assemble into a polymer network by alternating crystallographic interfaces<sup>8,9</sup>. The structure of DCV is symmetrical and has the greatest chemical complementarity when docked across the dimer interface reported previously<sup>7</sup> (Fig. 3a). In a model based on chemical shape matching and resistance mapping, DCV binds along the dimer ridge (defined by Y93, Q62, F37, F37', Q62' and Y93') and between and under the 28–35 loops to disrupt the natural packing of the NS5A N termini. The complementary effect of DCV and Syn-395 requires communication between two or more binding sites that cooperatively inhibit NS5A function. In this case, the dimer model was incorporated into an NS5A helical polymer (Fig. 3b) with the NS5A N termini extending from the helix axis and radiating out along the helical groove to position the NS4B/NS5A cleavage sites near the surface. Within the uninhibited NS5A polymer model, the hydrophobic residues from the N termini (28–35) pack along the floor of the helical axis defined by F37, Q62 and Y93. As in the dimer model, DCV binds along the NS5A helical axis, under and between the 28–35 loops, to disrupt the natural packing. However, in the NS5A polymer model, the inhibitor-induced perturbation affects the NS5A dimer in which DCV is bound and is also transmitted along the helical axis through P29–P35 loop interactions (Fig. 3b) to inhibit multiple NS5A proteins. Resistance mutations, which are typically smaller amino acids (L31V, Y93H), partially compensate for the DCV-induced disturbance and are able to restore HCV replication function in the presence of bound inhibitor(s). Syn-395 binding, either adjacent to or a few dimers to either side of DCV, potentiates the effect of DCV to introduce a conformational change that resensitizes the resistant NS5A towards inhibition (Fig. 3c).

Consistent with this hypothesis, only specific pairs of NS5A compounds exhibited this unparalleled synergistic effect. Synergy was not observed for BMS-313, a diastereomer of DCV (Extended Data Fig. 1) that is inactive against wild-type and resistant GT-1a replicons (Extended Data Fig. 9) at Syn-395 concentrations  $\leq 40$  nM.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 19 March; accepted 9 September 2015.**

**Published online 4 November 2015.**

1. Mohd Hanafiah, K., Groeger, J., Flaxman, A. D. & Wiersma, S. T. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* **57**, 1333–1342 (2013).
2. Shepard, C. W., Finelli, L. & Alter, M. J. Global epidemiology of hepatitis C virus infection. *Lancet Infect. Dis.* **5**, 558–567 (2005).

3. Pawlotsky, J. M. NS5A inhibitors in the treatment of hepatitis C. *J. Hepatol.* **59**, 375–382 (2013).
4. Guedj, J. *et al.* Modeling shows that the NS5A inhibitor daclatasvir has two modes of action and yields a shorter estimate of the hepatitis C virus half-life. *Proc. Natl Acad. Sci. USA* **110**, 3991–3996 (2013).
5. Macdonald, A. & Harris, M. Hepatitis C virus NS5A: tales of a promiscuous protein. *J. Gen. Virol.* **85**, 2485–2502 (2004).
6. He, Y., Staschke, K. A. & Tan, S.-L. in *Hepatitis C Viruses Genomes and Molecular Biology* (ed. Tan, S.-L.) Ch. 9, 267–291 (Horizon Bioscience, 2006).
7. Tellinghuisen, T. L., Marcotrigiano, J. & Rice, C. M. Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature* **435**, 374–379 (2005).
8. Love, R. A., Brodsky, O., Hickey, M. J., Wells, P. A. & Cronin, C. N. Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J. Virol.* **83**, 4395–4403 (2009).
9. Lambert, S. M. *et al.* The crystal structure of NS5A domain 1 from genotype 1a reveals new clues to the mechanism of action for dimeric HCV inhibitors. *Protein Sci.* **23**, 723–734 (2014).
10. Gao, M. *et al.* Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature* **465**, 96–100 (2010).
11. Quinkert, D., Bartenschlager, R. & Lohmann, V. Quantitative analysis of the hepatitis C virus replication complex. *J. Virol.* **79**, 13594–13605 (2005).
12. Pietschmann, T., Lohmann, V., Rutter, G., Kurpanek, K. & Bartenschlager, R. Characterization of cell lines carrying self-replicating hepatitis C virus RNAs. *J. Virol.* **75**, 1252–1264 (2001).
13. O'Boyle, D. R. II *et al.* Characterizations of HCV NS5A replication complex inhibitors. *Virology* **444**, 343–354 (2013).
14. Fridell, R. A., Qiu, D., Wang, C., Valera, L. & Gao, M. Resistance analysis of the hepatitis C virus NS5A inhibitor BMS-790052 in an *in vitro* replicon system. *Antimicrob. Agents Chemother.* **54**, 3641–3650 (2010).
15. Pelosi, L. A., Voss, S., Liu, M., Gao, M. & Lemm, J. A. Effect on hepatitis C virus replication of combinations of direct-acting antivirals, including NS5A inhibitor daclatasvir. *Antimicrob. Agents Chemother.* **56**, 5230–5239 (2012).
16. Nettles, R. E. *et al.* BMS-824393 is a potent HCV NS5A inhibitor with substantial antiviral activity when given as monotherapy in subjects with chronic G1 HCV infection. *61th Ann. Meeting AASLD* (2010).
17. Sandgren, E. P. *et al.* Complete hepatic regeneration after somatic deletion of an albumin-plasminogen activator transgene. *Cell* **66**, 245–256 (1991).
18. Hernandez, D., Zhou, N., Ueland, J., Monikowski, A. & McPhee, F. Natural prevalence of NS5A polymorphisms in subjects infected with hepatitis C virus genotype 3 and their effects on the antiviral activity of NS5A inhibitors. *J. Clin. Virol.* **57**, 13–18 (2013).
19. McCarville, J. F., Seifer, M., Standring, D. N., Mayers, D. L. & IDX-06A-001 Investigator Team. Treatment-emergent variants following 3 days of monotherapy with IDX719, a potent, pan-genotypic NS5A inhibitor, in subjects infected with HCV genotypes 1–4. *J. Hepatol.* **58** (suppl. 1), S491–S492 (2013).
20. Everson, G. T. *et al.* Phase 2b study of the interferon-free and ribavirin-free combination of daclatasvir, asunaprevir, and BMS-791325 for 12 weeks in treatment-naïve patients with chronic HCV genotype 1 infection. *64th Ann. Meeting AASLD* (2013).
21. Penin, F. *et al.* Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J. Biol. Chem.* **279**, 40835–40843 (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Pizzano for conducting *in vivo* experiments and K. Snow for bioanalytical support.

**Author Contributions** J.-H.S. and M.G. initiated synergy combination experiments. J.-H.S., D.R.O., R.A.F., C.W., S.B.R., P.N., M.B., Y.-K.W., M.L., K.R., J.A.L. and M.G. designed and performed the replicon screen for identifying synergy compounds, experiments with infectious HCV, genotype coverage, colony elimination, inhibitor-binding, isolation and mapping of resistant variants, and *in vitro* combination studies. M.G., M.B., J.K., N.A.M. and M.C. designed the overall virology studies and provided input to the overall research direction. D.R.L. constructed models of HCV NS5A and D.R.L., N.A.M. and M.G. contributed to the development of mechanistic hypotheses. B.M.J., F.M., M.J.N., M.B. and M.K. designed and interpreted *in vivo* studies. Y.T., P.H., M.B. and N.A.M. designed and synthesized the discussed compounds.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G. (Min.Gao@bms.com).

## METHODS

**NS5A compounds.** All compounds used in these studies were prepared at Bristol-Myers Squibb (BMS). Synthetic routes and analytical characterizations for BMS compounds are provided (Supplementary Information)<sup>22–26</sup>. LDV was prepared as described previously<sup>27</sup>.

**Replicon cell lines and assays.** All HCV replicon cell lines were isolated and characterized at BMS using a Huh7 cell line obtained from R. Bartenschlager. Each cell line tested negative for mycoplasma. Replicon assays for EC<sub>50</sub> determination<sup>14</sup>, colony elimination for the determination of resistance barrier<sup>15</sup>, and generation of HCV chimaeric replicon cell lines<sup>16</sup> have been described previously.

**Synergy effect assay.** The standard HCV replicon assay was described previously<sup>12</sup>. A two-way titration was used to determine the effectiveness of two NS5A compounds with the potential for synergy: the potency of DCV was measured by titration of DCV in the absence or presence of increasing concentrations of Syn-395 in HCV replicon cell lines, and the potency of Syn-395 was measured by titration of Syn-395 in the absence or presence of increasing concentrations of DCV. An HCV NS3 protease inhibitor and/or NS5B polymerase inhibitor were used as controls.

**Infectious HCV and assay.** A GT-2a JFH chimaeric virus with structural genes derived from the J6CF strain has been described<sup>28</sup>. A chimaeric virus with NS5A sequences derived from GT-1a was made by subcloning a BstE1 fragment in the J6-JFH virus with a BstE1 fragment amplified by PCR from an H77 clone. The resulting clone encodes an NS5A protein with amino acids 1–427 derived from H77c and amino acids 428–466 derived from JFH. Media was collected from JFH1 virus RNA-transfected cultures and used to infect Huh-7.5 cells plated in T175 flasks as previously described<sup>28</sup>. HCV replication was detected using Luc activity and microscopic visualization of cytopathic effects. Infected cells were split 1 to 4 and counted until the amount of Luc signal obtained was  $\sim 2 \times 10^6$  units; then 1 ml aliquots of cells were added to separate T25 flasks containing the appropriate amount of inhibitors or DMSO. The individual T25 flasks were then trypsinized every 2 or 3 days and resuspended in 4 ml complete media with 1 ml passed to a new flask and 1 ml used for Luc monitoring or RNA isolation. DCV rebounded virus was maintained in DCV until two passages before reaching baseline ( $\sim 2 \times 10^6$  Luc units ml<sup>-1</sup>). Isolated RNA was used for RT-PCR, qRT-PCR and/or sequencing. Measurement from qRT-PCR and Luc provided estimates of  $1.04 \times 10^{-2}$  Luc units per genome.

**PXB mouse study.** The study was performed under contract by PhoenixBio; use of animals was approved by the Animal Ethics Committee of PhoenixBio. Investigators at PhoenixBio were blinded to the identity of the compounds. Studies were performed in male PXB mice: uPA<sup>+/+</sup>/SCID [uPA<sup>+/+</sup>; B6SJL-TgN(Alb1Plau) 144Bri, SCID: C.B-17/1cr-scld/scld Jcl]; these mice are homozygous for the uPA transgene and possess the SCID trait. Mice had human hepatocytes transplanted to the liver. Ages ranged from 10–14 weeks. The study objective was to examine the anti-HCV effects of BMS-393 and Syn-776, both singly and in combination, in HCV-1a-infected PXB mice.

A tolerability study (14 days, QD,  $n = 5$ ) of vehicle and compounds at two doses (BMS-393 plus Syn-776 at  $0.4 + 5 \text{ mg kg}^{-1}$  and  $4 + 25 \text{ mg kg}^{-1}$ ) and a pharmacokinetic study (24 h) at a single dose (BMS-393 plus Syn-776 at  $0.4 + 5 \text{ mg kg}^{-1}$ , respectively) were performed in uninfected mice with low human albumin expression ( $< 70\%$  huAlb). PXB mice ( $n = 18$ ) used for the efficacy study had an estimated liver replacement index  $\geq 70\%$  based on the blood concentration of human albumin (huAlb  $> 70\%$ ). Human albumin expression was monitored throughout the experiment. Mice were chronically infected with HCV GT-1a and had HCV titres  $\geq 1.0 \times 10^6$  copies ml<sup>-1</sup> at the week before group assignment (actual levels on day  $-7$ :  $2.3 \times 10^7$  to  $3.6 \times 10^8$  copies ml<sup>-1</sup>). Primers and probes used for HCV RNA analysis are as follows. Forward primer (nucleotides 130–146), 5'-CGGGAGAGCCATAGTGG-3'; reverse primer (nucleotides 290–272), 5'-AGTACCACAAGGCCTTTCG-3'. TaqMan probe (nucleotides 148–168), 5'-CTGCGGAACCGGTGAGTACAC-3' (dye: 6-carboxyfluorescein (FAM) for

5', 6-carboxyfluorescein (TAMRA) for 3'). The lowest quantification limit of the assay was  $4.0 \times 10^4$  copies ml<sup>-1</sup>.

**RT-PCR, DNA sequencing and analysis.** Viral RNA isolation from mouse serum and RT-PCR was performed as described previously<sup>10</sup>. Three sets of primer pairs were used to generate products for population and clonal DNA sequencing: (1) 5'-GCATTCAAGATCATGAGCGGTGAGGTCCC-3' and 5'-CTTCCTCTACGGATAGCAAGTTAGCCTTCAC-3'; (2) 5'-CCTGGAGCCCTTGATGTCGTGTGGTCTG-3' and 5'-CCAGGTCGGGGAACACGATGAGACGAGCTG G-3'; and (3) 5'-AGGGGCGAGTGCATGGATGAACCGGCTAATAGCCTT-3' and 5'-ACGTCCTGGTAATGGCTGTCCAGAAC-3'. Analysis of sequences was performed using Sequencher DNA program analysis (Gene Codes).

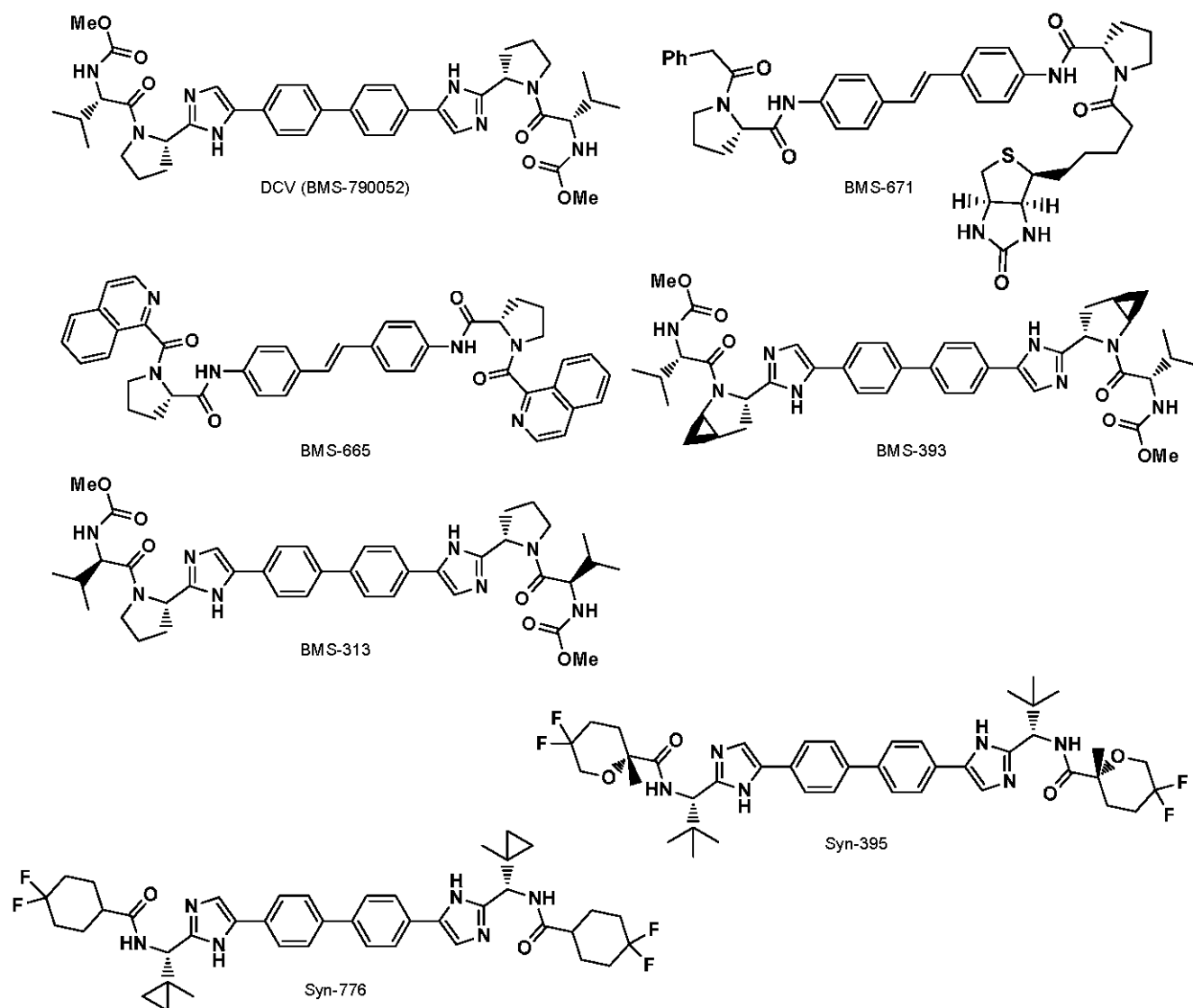
**NS5A domain I hexamer model.** Hexamer NS5A domain I models were constructed using X-ray and NMR structure fragments. Amino acids 36–194 were modelled based on the Tellinghuisen *et al.* dimer (PDB accession 1ZH1) structure; however, the side-chain conformation of the outward facing amino acids of the 1ZH1 dimer that make up the Love *et al.* dimer (PDB accession 3FQM) interface were modified to match the conformation observed in the 3FQM X-ray structure<sup>7,8</sup>. Note that the outward facing residues on the A-chain of 1ZH1 were conformationally matched with the B-chain of 3FQM while the B-chain of 1ZH1 was matched with the A-chain of 3FQM. By alternately overlaying the modified dimer onto the 3FQM structures and removing the 3FQM template, a hexameric NS5A polymer (ABABAB) was constructed that contained a crystallographic interface between each monomer.

**Docking DCV.** DCV was hand docked by aligning the long axis of the DCV di-imidazol-biphenyl core with the helical axis of the NS5A domain I hexamer model and sliding it along the deep trough of the polymer axis to maximize symmetrical, chemical and shape complementarities.

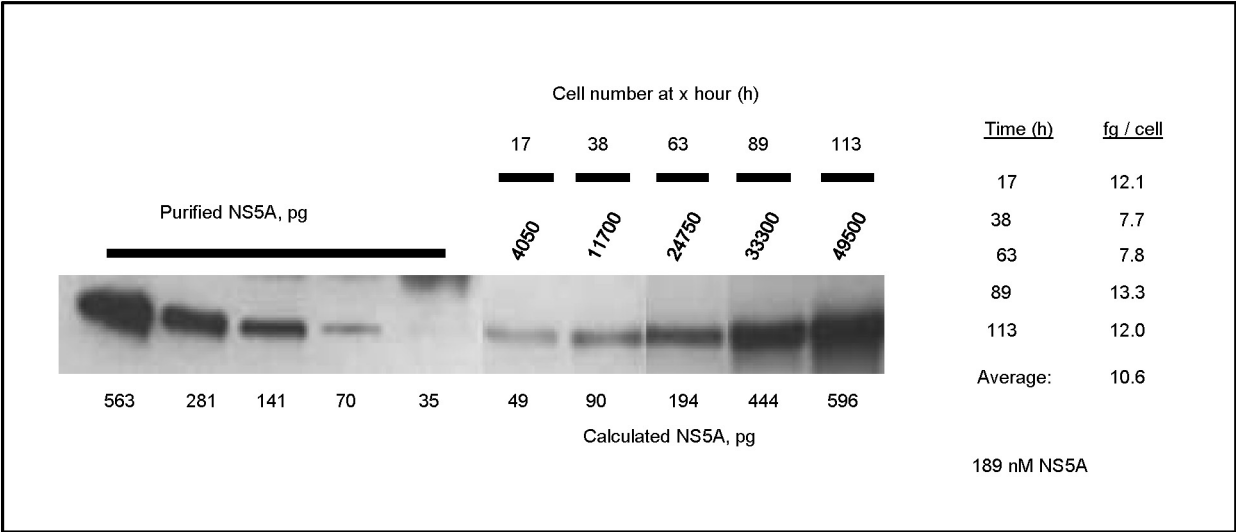
**Building in the N termini amphipathic  $\alpha$ -helix.** Residues 1–31 were modelled based on the Penin *et al.* PDB accession 1R7C NMR structure<sup>9</sup>. The N termini helix was hand docked into the shallow trench that extends from the helical axis along the groove towards the surface of the helix. The 1R7C structure was docked into the trench using a distance constraint, to allow for insertion of the missing residues 32–35, and to maximize symmetrical, chemical and shape complementarity. Residues 32–35 were modelled around DCV to complete the structures. The crystallographic Zn<sup>2+</sup> ions were retained.

**DCV/NS5A domain I dimer model.** The DCV/NS5A domain I dimer model was extracted directly from the hexamer model described earlier. Amino acid side-chain adjustments, structure superimpositions and amino acid gap filling were done using the protein design package in QUANTA (QUANTA Modelling Environment, release 2006, Accelrys Software). The images for Fig. 3 were generated using Maestro (Schrödinger LLC).

22. Pack, S. K. *et al.* Hepatitis C virus inhibitors. US patent 9,006,455 B2 (2015).
23. St. Laurent, D. R. *et al.* HCV NS5A replication complex inhibitors. Part 4. Optimization for genotype 1a replicon inhibitory activity. *J. Med. Chem.* **57**, 1976–1994 (2014).
24. Seebach, D., Naef, R. & Calderari, G.  $\alpha$ -Alkylation of  $\alpha$ -heterosubstituted carboxylic acids without racemization. *Tetrahedron* **40**, 1313–1324 (1984).
25. Nagase, R., Oguni, Y., Misaki, T. & Tanabe, Y. Practical and robust method for the preparation of Seebach and Frater's chiral template, *cis*-2-substituted 5-methyl (or phenyl)-1,3-dioxolan-4-ones. *Synthesis* **22**, 3915–3917 (2006).
26. Reddy, L. R., Gupta, A. P. & Liu, Y. Asymmetric synthesis of  $\alpha$ -amino acids by reduction of *N*-tert-butanefulfonyl ketimine esters. *J. Org. Chem.* **76**, 3409–3415 (2011).
27. Link, J. O. *et al.* Discovery of ledipasvir (GS-5885): a potent, once-daily oral NS5A inhibitor for the treatment of hepatitis C virus infection. *J. Med. Chem.* **57**, 2033–2046 (2014).
28. Baldick, C. J. *et al.* A novel small molecule inhibitor of hepatitis C virus entry. *PLoS Pathog.* **6**, e1001086 (2010).



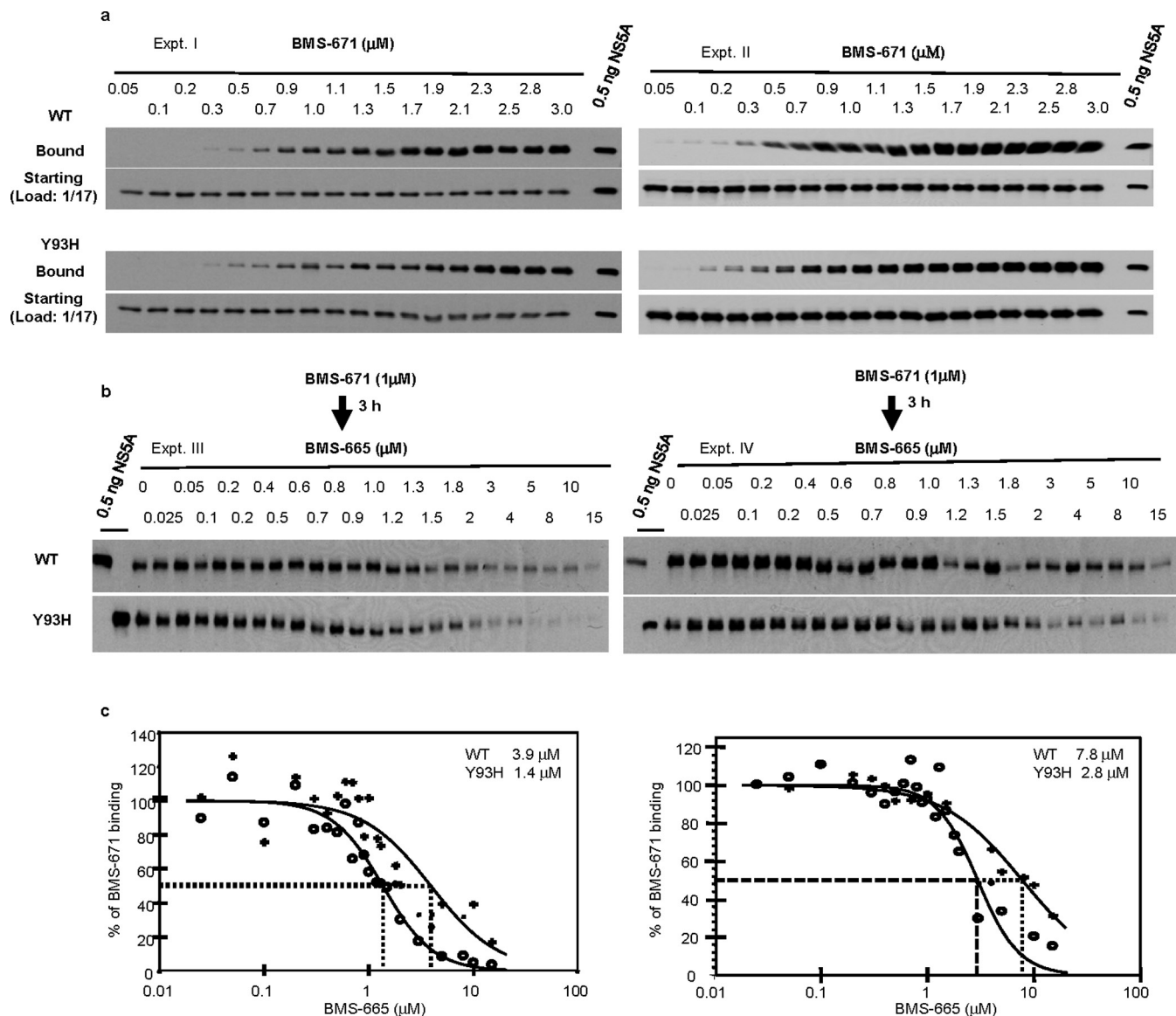
**Extended Data Figure 1 | Structures.** Structures of the Bristol-Myers Squibb (BMS) compounds used to characterize NS5A synergy: DCV (BMS-790052; also known as daclatasvir and Daklinza), BMS-313, BMS-665, BMS-393, BMS-671 (biotin-tagged), Syn-395 and Syn-776.



**Extended Data Figure 2 | Quantification of HCV NS5A in GT-1b (Con1) cells.** Cells were counted (4,050 to 49,500) and harvested at the indicated time points (17 to 113 h after seeding). The harvested cells were lysed and subjected to immunoblot analysis with polyclonal anti-NS5A sera<sup>13</sup>. NS5A in each sample was quantified using a standard curve derived

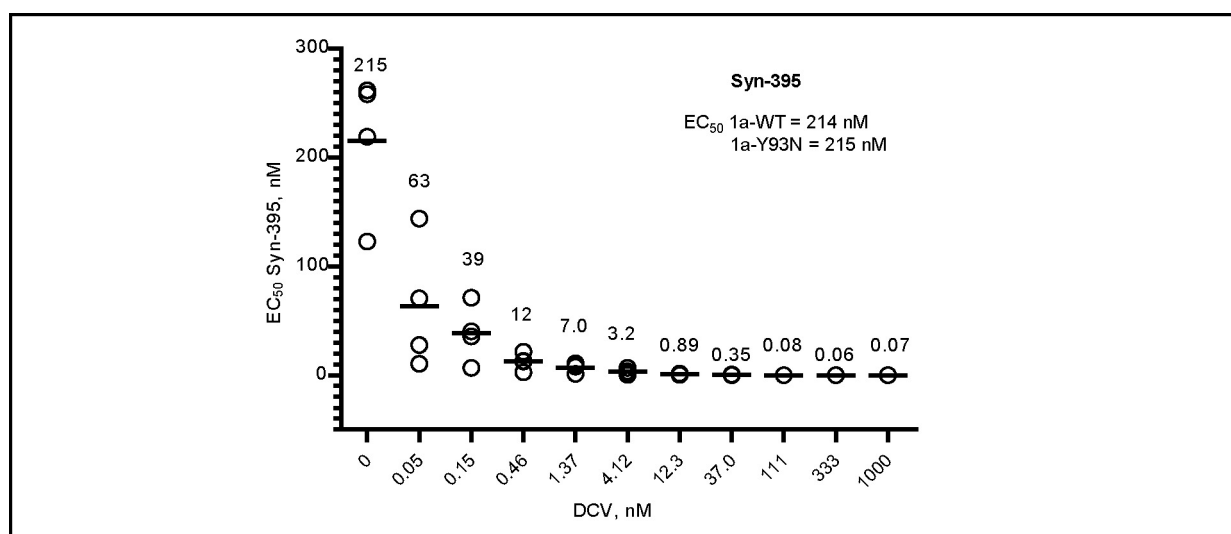
from purified NS5A protein. On the basis of the average concentration of NS5A in cells (189 nM) and the concentration of DCV at half-maximal inhibition ( $EC_{50}$  0.004 nM), the ratio of NS5A to DCV was 47,000:1. The immunoblot is representative of two independent experiments.



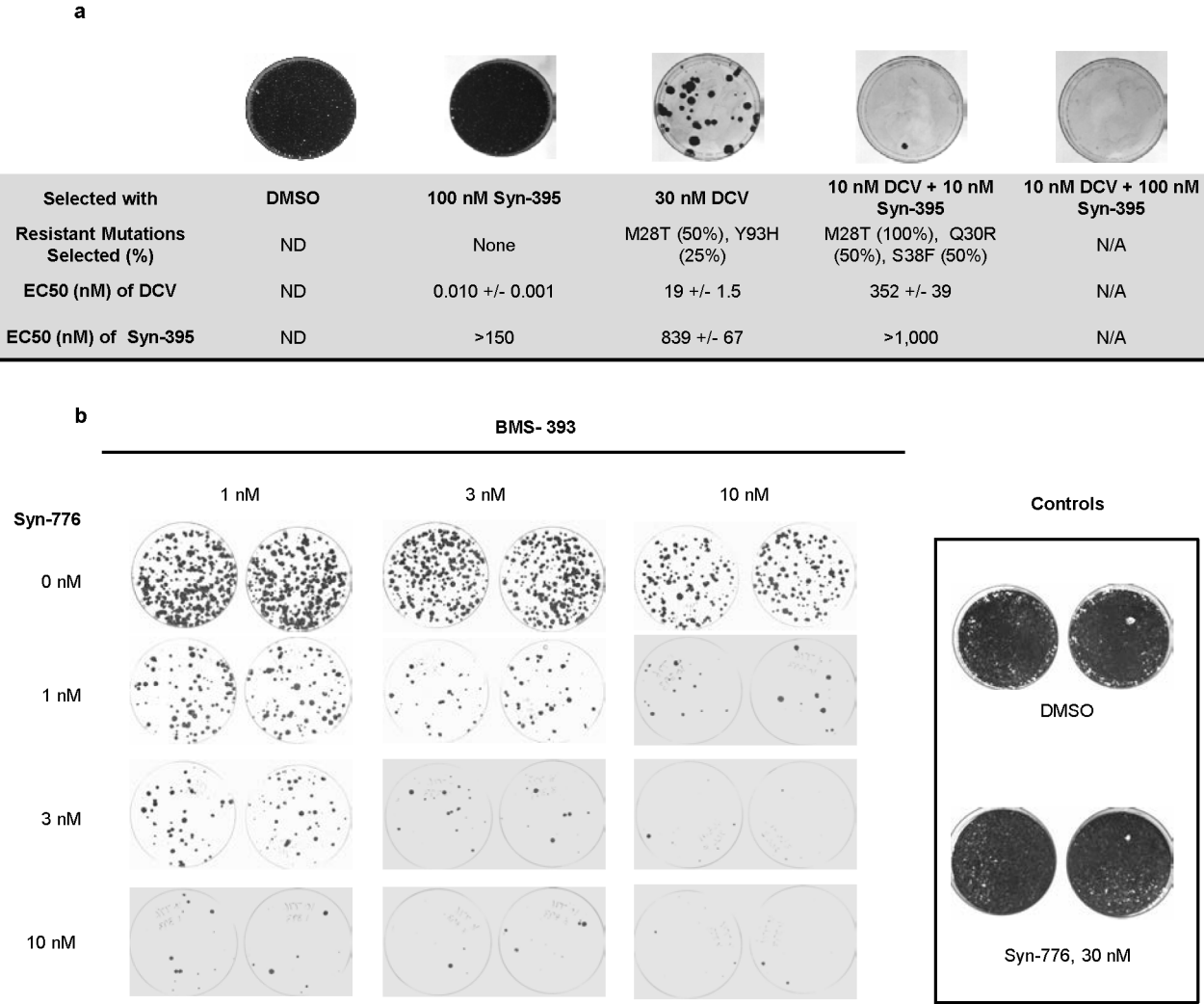


**Extended Data Figure 3 | Inhibition and compound binding to HCV NS5A.** **a**,  $\text{EC}_{50}$  values for BMS-671/BMS-665 on GT-1b wild-type and Y93H mutant are 33/22 nM and 7,800/1,800 nM, respectively. Increasing amounts of biotinylated inhibitor BMS-671 (0.05 to 3  $\mu\text{M}$ ) were added to wild-type (top) and Y93H resistant (bottom) GT-1b replicon cells for 16 h. Approximately 1 of 17 of the lysed cells were used as input controls and the remaining lysates were mixed with streptavidin-agarose beads. Proteins bound to streptavidin beads were fractionated on polyacrylamide gels and detected by immunoblot with polyclonal anti-NS5A sera<sup>13</sup>. Bound NS5A was quantified by scanning the signal intensity of each band.

Results from two independent experiments are shown. **b**, Western blot of NS5A wild type and Y93H bound to biotinylated BMS-671 in the presence of increasing concentrations of the competitor BMS-665. Proteins were detected and quantified as described earlier. Results from two independent experiments are shown. **c**, The relative intensity of NS5A in each sample shown in **b** expressed as a percentage of the signal from cells treated with BMS-671 alone. Values (wild type/Y93H = 3.9/1.4  $\mu\text{M}$  and 7.8/2.8  $\mu\text{M}$ ) represent the BMS-665 concentration required to reduce the binding of biotinylated BMS-671 to wild-type and Y93H NS5A (GT-1b) by 50%.

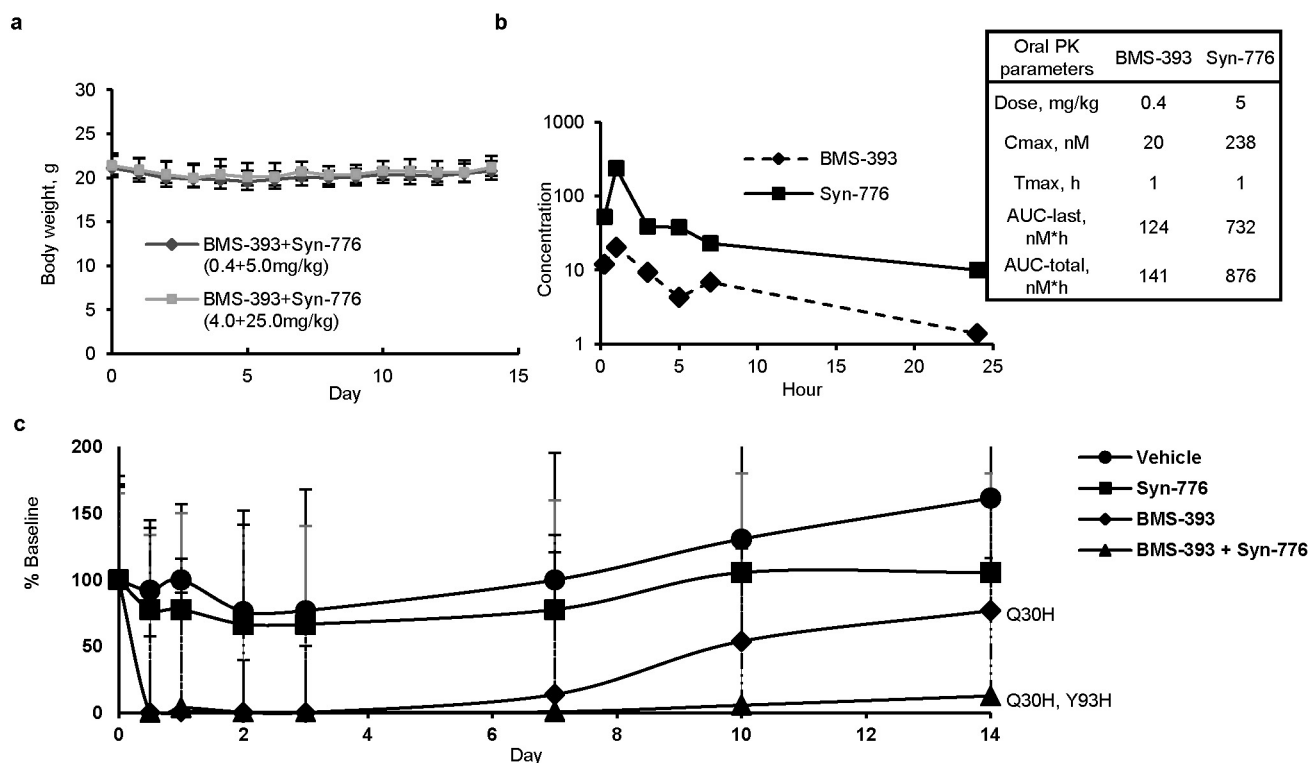


**Extended Data Figure 4 | Synergistic effect of DCV and Syn-395 on a GT-1a Y93N replicon.**  $EC_{50}$  values of Syn-395 were determined in the absence or presence of different concentrations of DCV. Individual values and the average of four independent experiments are shown.



**Extended Data Figure 5 | Colony elimination studies with GT-1a replicon cells. a,** The combination of DCV and Syn-395 enhances the resistance barrier compared to the resistance barrier of monotherapies<sup>14</sup>. GT-1a colonies surviving treatment for 7 days with DMSO (control), Syn-395 (100 nM), DCV (30 nM) or DCV (10 nM) plus Syn-395 (100 or 10 nM) were visualized after crystal violet staining. Surviving cells in duplicate plates (10 nM DCV plus 10 nM Syn-395 plus G418) were expanded for genotypic (resistance mapping) and phenotypic (EC<sub>50</sub> determination) analysis. No consensus substitutions were observed in Syn-395-treated cells. In DCV-selected cells, most variants contained a single amino acid substitution (M28T or Y93H) with lower level resistance

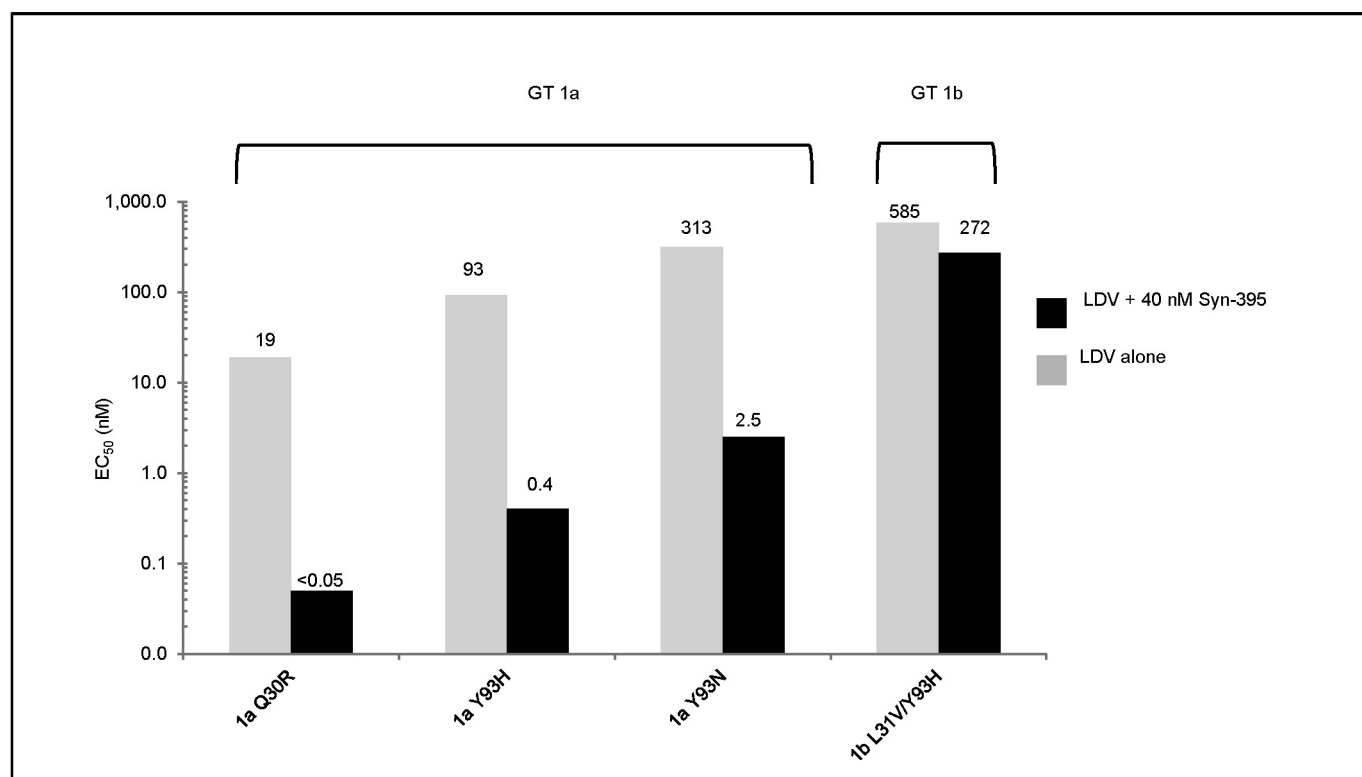
(EC<sub>50</sub> of 19 nM for DCV). In cells treated with the DCV plus Syn-395 combination, higher level resistance was observed (EC<sub>50</sub> of 352 nM for DCV), and linked substitutions containing M28T (M28T-Q30R, M28T-S38F or M28T-Q30R-S38F) were detected in 100% of the replicon cells. The pattern shown is representative of three independent experiments. **b,** Synergistic activity was confirmed with an alternative combination, BMS-393 and Syn-776. EC<sub>50</sub> values of BMS-393 (0.003 ± 0.002 nM) and Syn-776 (544 ± 179 nM) in GT-1a replicon cells provided a basis for selecting inhibitor concentrations for two-way titrations. The pattern shown is representative of two independent experiments.



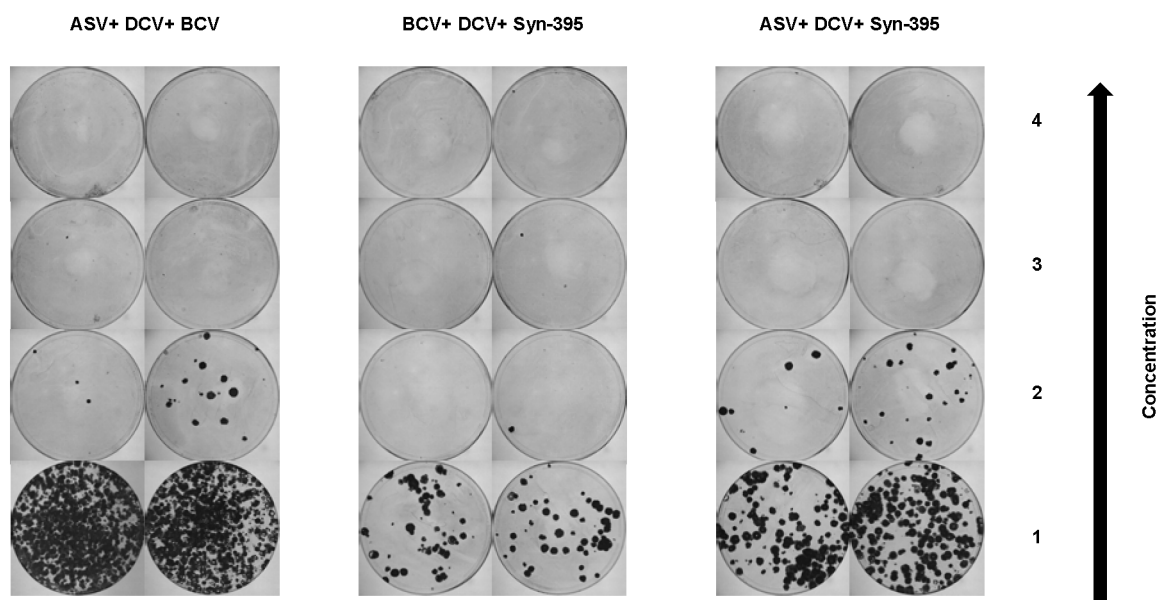
**Extended Data Figure 6 | PXB mouse study.** **a**, Body weights of PXB mice that had BMS-393 and Syn-776 administered concomitantly (0.4 + 5, and 4 + 25 mg kg<sup>-1</sup>) were stable during the in-life phase (90% or more compared to day 0). Data are presented as the mean and standard deviation (s.d.) for five mice. **b**, Serum concentration–time profiles of BMS-393 and Syn-776 (0.4 and 5 mg kg<sup>-1</sup>) administered concomitantly to PXB mice (day 1) are shown in the graph and pharmacokinetic (PK) parameter values are summarized in the adjacent box. To assess the systemic exposure to BMS-393 and Syn-776, pharmacokinetic parameters were calculated using non-compartmental methods. Each area under the curve (AUC) was calculated using the linear trapezoidal rule. Owing to small sample volumes, serum samples from three mice were pooled at each time point before analysis. Data are presented as the concentration (nM) of each pooled sample. **c**, Change in HCV titres for PXB mice dosed with vehicle only (filled circle), Syn-776 only (square; 15 mg kg<sup>-1</sup>), BMS-393 only (diamond; 0.4 mg kg<sup>-1</sup>) or a combination of BMS-393 plus Syn-776

(triangle; 0.4 + 15 mg kg<sup>-1</sup>) were plotted. Data for each group of mice (five) at each time point are presented as the percentage of baseline and the error bars indicate per cent variation at each time point. Mice dosed with either vehicle or Syn-776 had similar variations in titre, indicating no significant virus suppression. On day 14, 13% inhibition was observed for the group of mice treated with BMS-393 alone, versus 87% inhibition for the group treated with the BMS-393 plus Syn-776 combination. Genotypic analysis by population sequencing of day 14 samples revealed only wild-type sequences in samples derived from cohorts receiving vehicle or Syn-776 alone; a single Q30H substitution was predominant with BMS-393 treatment; a mixture of substitutions (mainly Q30H and Y93H) was observed with the combination of BMS-393 plus Syn-776. The genotypic analysis confirms that the BMS-393 plus Syn-776 combination confers a higher genetic barrier to the emergence of resistance than DCV alone, an *in vivo* validation of the synergistic effect.



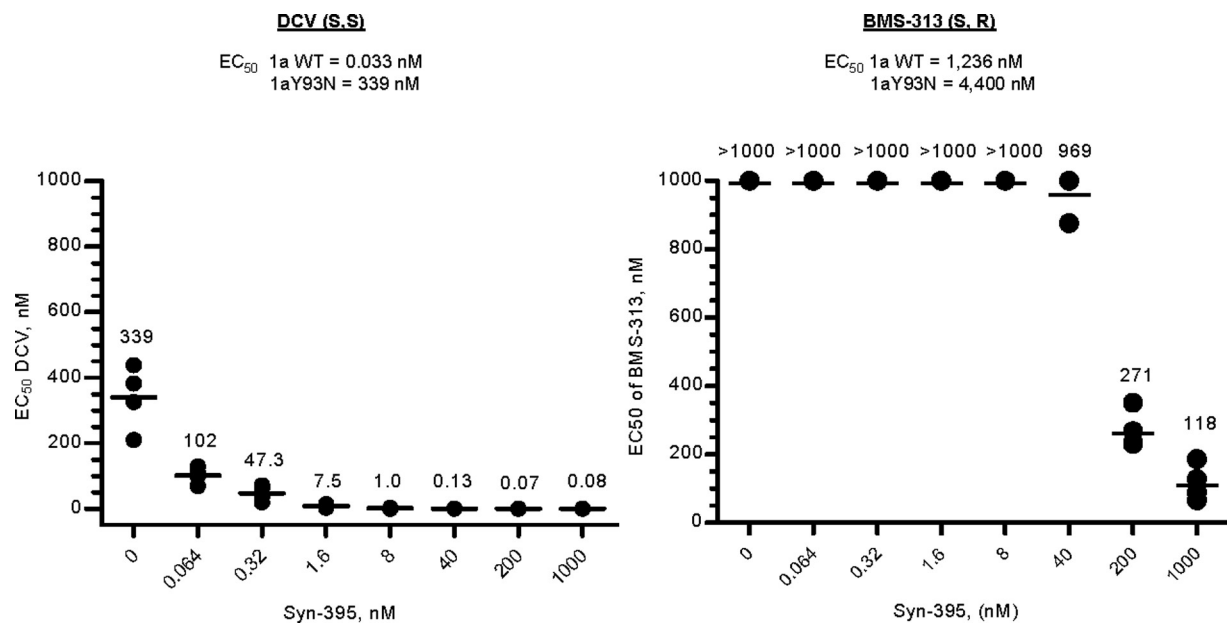


**Extended Data Figure 7 | Synergistic effect of LDV (GS-5885) and Syn-395.**  $EC_{50}$  values of LDV were determined by titrating LDV in the absence and presence of 40 nM Syn-395 using resistant replicon cells of different genotypes (GT-1a Q30R, Y93H, Y93N and GT-1b L31V–Y93H). Values are from a single experiment.



**Extended Data Figure 8 | Direct comparison of synergistic combinations versus DCV-3DAA.** A synergist (Syn-395;  $EC_{50}$  214 nM) was used to replace either an NS3 protease inhibitor (ASV;  $EC_{50}$  6 nM) or NS5B polymerase inhibitor (BCV;  $EC_{50}$  5 nM) in a three-drug combination to eliminate HCV replicon colonies *in vitro*. GT-1a replicon cells were treated with HCV inhibitors as indicated, and surviving colonies were visualized by crystal violet staining. Concentration (nM) of DAAs

increased as indicated by arrow: ASV + DCV + BCV (row 1: 6 + 0.04 + 5; row 2: 18 + 0.12 + 15; row 3: 60 + 0.4 + 50; row 4: 180 + 1.2 + 150); BCV + DCV + Syn-395 (row 1: 5 + 0.04 + 5; row 2: 15 + 0.12 + 15; row 3: 50 + 0.4 + 50; row 4: 150 + 1.3 + 150); and ASV + DCV + Syn-395 (row 1: 6 + 0.04 + 5; row 2: 18 + 0.12 + 15; row 3: 60 + 0.4 + 50; row 4: 180 + 1.2 + 150). The profile is representative of four independent experiments.



**Extended Data Figure 9 | Comparing synergy for a specific pair of NS5A compounds.** Experiments were performed on a GT-1a Y93N replicon as described in the legend of Fig. 1b. Unlike the synergy observed between DCV (SSSS stereochemistry) and BMS-395, no synergy was

observed at concentrations  $\leq 40$  nM when a diastereomer of DCV, BMS-313 (RSSR stereochemistry) was paired with BMS-395. Individual values and the average from four independent experiments are shown.

Extended Data Table 1 | Clonal analysis of HCV from PXB mice treated with BMS-393 and BMS-393 plus BMS-776

Day 14 (total number)	WT	M28	Q30							L31		Y93		Linkage (number of clones)
		T	E	H	K	G	R	P	D	V	M	H	N	
BMS-393 0.4 mg/kg (289)														P29S + Q30E (1)
														K26E + Q30H (1)
														Q30H + Y93H (1)
														L31V + Y93R (1)
														K24R + M28T (1)
														M28T + Q30R (1)
														M28V + Q30H (1)
% of total clones	0	27.6	6.2	44.5	3.1	0	0.3	0.3	0.3	12.5	3.1	0	0	2.1
BMS-393 + Syn-776 0.4 + 15 mg/kg (381)														M28T + Q30R + Y93H (1)
														M28T + Q30R (4)
														M28R + Y93N (1)
														M28T + Y93H (1)
														M28T + Y93C (2)
														M28T + Y93N (1)
														P29L + Q30H (1)
														Q30K + Y93R (1)
														Q30K + Y93H (2)
														Q30H + L31P (2)
														Q30H + L31V (1)
														Q30H + Y93H (1)
														K26R + Q30E (1)
														Q30H + Y93N (1)
														Q30K + L31M (1)
														L31M + Y93H (4)
														L31M + Y93N (1)
% of total clones	6	10	8.1	17.6	15.7	0.3	0	0	0	2.4	10.2	14.2	8.4	7.1

A single Q30H substitution was predominant with BMS-393 selection (44.5% of 289 clones); linked substitutions were more prevalent with the BMS-393 plus Syn-776 combination (7.1% of 381 clones) compared with BMS-393 alone (2.1% of 289 clones). The presence of wild-type (WT) virus in mice treated with the BMS-393 plus Syn-776 combination suggests that some DCV-resistant variants with a single substitution are suppressed as effectively as wild type by the treatment.



# Epigenetic silencing of T<sub>H</sub>1-type chemokines shapes tumour immunity and immunotherapy

Dongjun Peng<sup>1</sup>, Ilona Kryczek<sup>1,2</sup>, Nisha Nagarsheth<sup>1,2</sup>, Lili Zhao<sup>3</sup>, Shuang Wei<sup>1</sup>, Weimin Wang<sup>1</sup>, Yuqing Sun<sup>4</sup>, Ende Zhao<sup>1</sup>, Linda Vatan<sup>1</sup>, Wojciech Szeliga<sup>1</sup>, Jan Kotarski<sup>5</sup>, Rafał Tarkowski<sup>5</sup>, Yali Dou<sup>4</sup>, Kathleen Cho<sup>4,6</sup>, Sharon Hensley-Alford<sup>7</sup>, Adnan Munkarah<sup>7</sup>, Rebecca Liu<sup>6,8</sup> & Weiping Zou<sup>1,2,6,9</sup>

**Epigenetic silencing including histone modifications and DNA methylation is an important tumorigenic mechanism<sup>1</sup>. However, its role in cancer immunopathology and immunotherapy is poorly understood. Using human ovarian cancers as our model, here we show that enhancer of zeste homologue 2 (EZH2)-mediated histone H3 lysine 27 trimethylation (H3K27me3) and DNA methyltransferase 1 (DNMT1)-mediated DNA methylation repress the tumour production of T helper 1 (T<sub>H</sub>1)-type chemokines CXCL9 and CXCL10, and subsequently determine effector T-cell trafficking to the tumour microenvironment. Treatment with epigenetic modulators removes the repression and increases effector T-cell tumour infiltration, slows down tumour progression, and improves the therapeutic efficacy of programmed death-ligand 1 (PD-L1; also known as B7-H1) checkpoint blockade<sup>2–4</sup> and adoptive T-cell transfusion<sup>5</sup> in tumour-bearing mice. Moreover, tumour EZH2 and DNMT1 are negatively associated with tumour-infiltrating CD8<sup>+</sup> T cells and patient outcome. Thus, epigenetic silencing of T<sub>H</sub>1-type chemokines is a novel immune-evasion mechanism of tumours. Selective epigenetic reprogramming alters the T-cell landscape<sup>6</sup> in cancer and may enhance the clinical efficacy of cancer therapy.**

Cancer immunotherapy has demonstrated therapeutic responses<sup>4,5,7,8</sup>. Yet, the objective responses have been manifested in a fraction of patients. We hypothesized that immune-protective signature genes might be epigenetically silenced in cancer and in turn affect cancer progression and clinical responses to immunotherapy. Cancer epigenetic silencing often includes EZH2-mediated histone modifications and DNMT-mediated DNA methylation. DZNep<sup>9</sup> is an inhibitor of all S-adenosyl-methionine (SAM)-dependent enzymes including EZH2, and EPZ6438<sup>10</sup> may specifically inhibit EZH2. 5-aza-2'-deoxycytidine (5-AZA-dC) is a DNMT inhibitor. Hence, we used these agents to reprogram epigenetic pathways and to test our hypothesis.

In the first setting, we established ID8 ovarian cancer in C57/BL6 mice and treated them with low doses of DZNep and EPZ6438, 5-AZA-dC and their combination. Treatment with single agent had minimal effects on tumour volume, whereas the combinatorial treatment caused tumour reduction, increased tumour infiltrating T cells and T<sub>H</sub>1-type chemokine expression (Extended Data Fig. 1a–e). We observed similar tumour volume (Extended Data Fig. 1f) in ID8-bearing, female NOD-scid Il2r<sup>null</sup> (NSG) mice<sup>11</sup>, that received identical treatment. Thus, epigenetic reprogramming elicits potent tumour immunity and blocks cancer progression.

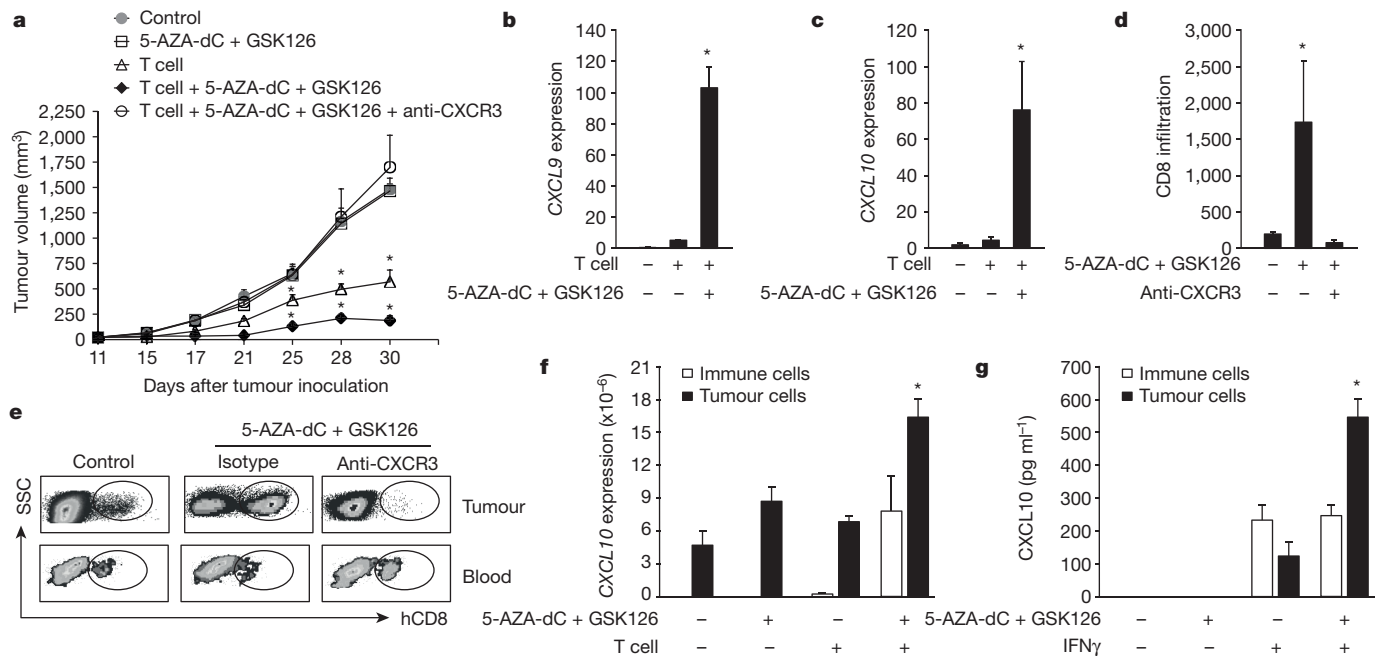
We have demonstrated the relevance of the inhibitory B7-H1 (PD-L1) signalling blockade in human cancer<sup>2,3</sup>. In the second setting, we tested the role of DZNep and 5-AZA-dC in anti-tumour immunity elicited by PD-L1 blockade. We observed reduced tumour volume in mice treated with anti-PD-L1 or DZNep plus 5-AZA-dC. The combination

reduced tumour volume, increased tumour-infiltrating CD8<sup>+</sup> T cells and T<sub>H</sub>1-type chemokine expression (Extended Data Fig. 1g–i). Thus, changes in the epigenetic program can augment the therapeutic efficacy of PD-L1 blockade therapy.

In the third setting, we examined the effects of these two epigenetic modulators on adoptive T-cell therapy in our NSG model. We established human ovarian cancer in NSG mice and generated tumour-associated antigen (TAA)-specific CD8<sup>+</sup> T cells<sup>12–14</sup>. Then, the NSG mice were treated with DZNep and 5-AZA-dC, and/or TAA-specific CD8<sup>+</sup> T cells. In the absence of TAA-specific T cells, treatment with DZNep and 5-AZA-dC had minimal effects on tumour progression. TAA-specific CD8<sup>+</sup> T cells reduced tumour volume (Extended Data Fig. 1j). Consequently, treatment with DZNep and 5-AZA-dC improved the therapeutic efficacy of T-cell therapy, and elevated T<sub>H</sub>1-type chemokine expression, and tumour effector T-cell infiltration (Extended Data Fig. 1j–l). GSK126 is a selective inhibitor of EZH2 methyltransferase activity<sup>15</sup>. We used GSK126 in the human T-cell therapy setting. Treatment with low doses of GSK126 and/or 5-AZA-dC had no effect on tumour growth in the absence of T-cell transfection (Extended Data Fig. 1m). Treatment with GSK126 and 5-AZA-dC synergistically improved the therapeutic efficacy of T-cell therapy (Fig. 1a and Extended Data Fig. 1m), increased tumour CXCL9 and CXCL10 expression (Fig. 1b, c), and CD8<sup>+</sup> T-cell infiltration (Fig. 1d, e and Extended Data Fig. 1n), and had minimal effects on TNFα and IFNγ expression (Extended Data Fig. 1o–q). Administration of anti-human CXCR3 abrogated the role of GSK126 and 5-AZA-dC treatment in tumour progression (Fig. 1a), and blocked T-cell tumour trafficking (Fig. 1d, e), and had no effect on T-cell apoptosis (Extended Data Fig. 1r, s). CXCR3<sup>+</sup>CD8<sup>+</sup> T cells were observed in the spleen (Extended Data Fig. 1t) and blood (Fig. 1e) with or without anti-CXCR3 treatment. To determine the major source of T<sub>H</sub>1-type chemokines, we isolated human tumour and immune cells from tumour tissues in the NSG mice (Fig. 1a–e). We found that administration of GSK126 and 5-AZA-dC increased tumour CXCL10 expression with or without T-cell transfection (Fig. 1f). Regardless of the treatment, the levels of CXCL10 expression were higher in tumour than immune cells (Fig. 1f). We extended our studies to tumour and immune cells from ovarian cancer patients. Although EZH2 may regulate naive CD4<sup>+</sup> T-cell interferon-γ (IFNγ) expression<sup>16</sup>, GSK126 and 5-AZA-dC treatment increased CXCL10 production by tumour, not by T cells (Fig. 1g), and tumour cells produced higher levels of CXCL10 than immune cells (Fig. 1g). Thus, epigenetic reprogramming may predominantly target tumour T<sub>H</sub>1-type chemokine expression.

Next, we investigated how epigenetic modulators regulate tumour T<sub>H</sub>1-type chemokine expression. We initially examined the effect of DZNep on T<sub>H</sub>1-type chemokine expression in primary human ovarian

<sup>1</sup>Department of Surgery, University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA. <sup>2</sup>Graduate Program in Immunology, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>3</sup>Department of Biostatistics, University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA. <sup>4</sup>Department of Pathology, University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA. <sup>5</sup>The First Department of Gynecologic Oncology and Gynecology, Medical University in Lublin, Lublin 20-081, Poland. <sup>6</sup>The University of Michigan Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>7</sup>Department of Women's Health Services, Henry Ford Health System, Detroit, Michigan 48202, USA. <sup>8</sup>Department of Obstetrics and Gynecology, University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA. <sup>9</sup>Graduate Program in Tumor Biology, University of Michigan, Ann Arbor, Michigan 48109, USA.



**Figure 1 | Epigenetic reprogramming alters cancer T-cell immunotherapy.** Effects of GSK126 and 5-AZA-dC on T-cell immunotherapy. Ovarian-cancer-bearing NSG mice were treated with GSK126, 5-AZA-dC and/or anti-CXCR3, and received TAA-specific CD8<sup>+</sup> T-cell transfusion. **a–f**, Tumour volume (**a**), tumour chemokine expression (**b**, **c**) and numbers of tumour-infiltrating T cells per gram of tissue (**d**) and intratumoural and blood T cells (**e**) were

shown. SSC, side scatter; hCD8, human CD8<sup>+</sup> T cell. **f**, *CXCL10* transcripts were quantified in tumour and immune cells from tumour tissues in NSG model and normalized to *GAPDH* mRNA level. **g**, Tumour and immune cells ( $1 \times 10^6$  per ml) were isolated from ovarian cancer from patients, and were treated for 48 h as indicated. CXCL10 protein was measured by ELISA (mean  $\pm$  s.d.,  $n = 5$ , \* $P < 0.05$ , Mann–Whitney *U*-test).

cancer cells in response to IFN $\gamma$  treatment. DZNep promoted CXCL9 and CXCL10 mRNA and protein expression (Extended Data Fig. 2a–c), and had no effect on IFN $\gamma$  receptor (IFNGR2) (Extended Data Fig. 2d). We tested two inhibitors of histone H3 lysine 9 methyltransferase G9a, BIX01294 and UNC0638. The inhibition of G9a had no effect on T<sub>H</sub>1-type chemokine expression (Extended Data Fig. 2e). Thus, methylation of H3K27 may mediate T<sub>H</sub>1-type chemokine repression in cancer.

EZH2 is the catalytic subunit of the H3K27 methyltransferase complex<sup>9,17</sup>. DZNep treatment reduced EZH2 protein expression and H3K27 trimethylation (H3K27me3) in primary ovarian cancer cells (Extended Data Fig. 2f). Similarly, EPZ6438 reduced H3K27 trimethylation and increased CXCL9 expression in mouse ID8 cells (Extended Data Fig. 2g, h). We genetically knocked down human EZH2 expression with lentivirus-based short hairpin RNA targeting *EZH2* (shEZH2). shEZH2 reduced the expression of EZH2 and H3K27me3 (Extended Data Fig. 2i), and resulted in elevated T<sub>H</sub>1-type chemokine mRNA and protein expression (Fig. 2a, b), and no change in IFNGR2 and HLA-B (Extended Data Fig. 2j, k). Thus, EZH2 and its histone methyltransferase activity mediate T<sub>H</sub>1-type chemokine repression in primary ovarian cancer cells.

We tested whether EZH2-mediated T<sub>H</sub>1-type chemokine repression depends on H3K27me3 changes at the promoter level. Chromatin immunoprecipitation (ChIP) assays revealed that IFN $\gamma$  treatment reduced H3K27me3 levels on the promoters of CXCL9 and CXCL10 (Extended Data Fig. 2l, m). DZNep (Extended Data Fig. 2l, m) and shEZH2 (Fig. 2c) largely removed H3K27me3 on the promoters of CXCL9 and CXCL10, and subsequently increased IFN $\gamma$ -induced chemokine gene expression (Fig. 2a, b). As a positive control, shEZH2 reduced H3K27me3 occupancy on the promoter of the *HOXB1* gene (Extended Data Fig. 2n)<sup>18</sup>. Thus, H3K27me3 removal results in the abrogation of the T<sub>H</sub>1-type chemokine gene silencing in primary ovarian cancer.

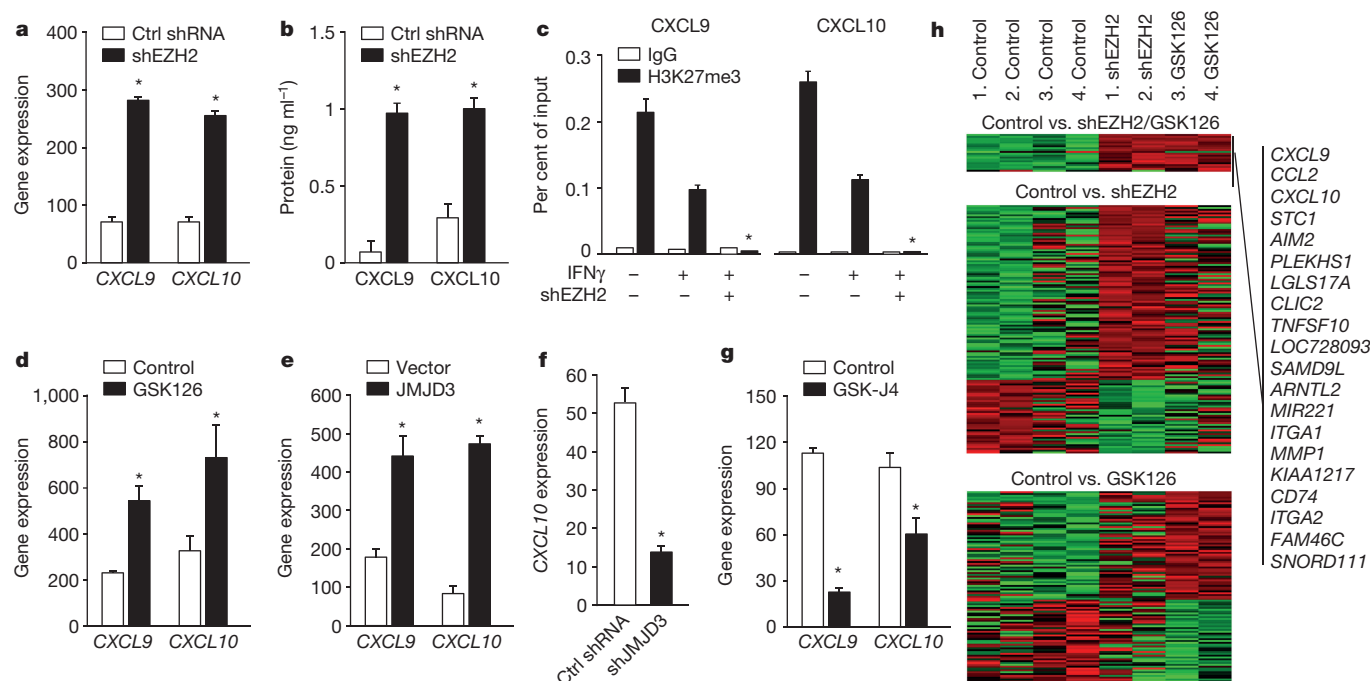
GSK126 is a selective inhibitor of EZH2 methyltransferase activity<sup>15</sup>. GSK126 treatment abolished the global level of H3K27me3 without inhibiting EZH2 (Extended Data Fig. 3a, b), *IFNGR2* expression or cell survival (Extended Data Fig. 3c, d). GSK126 treatment led to higher levels of IFN $\gamma$ -induced T<sub>H</sub>1-type chemokine expression in two

primary and three established ovarian cancer cell lines (Fig. 2d and Extended Data Fig. 3e–h). JmjC-domain-containing protein (JMJD3) is an H3K27-specific demethylase<sup>18</sup>. Ectopic expression of JMJD3 reduced H3K27me3 (Extended Data Fig. 3i), increased T<sub>H</sub>1-type chemokine expression (Fig. 2e and Extended Data Fig. 3j), and had no effect on *HLA-B* and *IFNGR2* expression (Extended Data Fig. 3k, l). JMJD3 deficiency introduced by a shJMJD3 inhibited CXCL10 (Fig. 2f), but not *IFNGR2* expression (Extended Data Fig. 3m).

GSK-J4, a catalytic site inhibitor, may target members of the H3K27-specific JmjC demethylase subfamily<sup>19</sup>. GSK-J4 treatment increased H3K27me3 (Extended Data Fig. 3n) and reduced CXCL9 and CXCL10 expression (Fig. 2g), and had no effects on H3K4me1, H3K4me2 or H3K4me3 or *IFNGR2* expression (Extended Data Fig. 3n, o).

To define the gene profile altered by EZH2 and H3K27me3 in response to IFN $\gamma$ , we performed several microarrays in primary ovarian cancer cells transfected with shEZH2, control, or treated with GSK126 and medium. We found that 155 and 124 genes were altered by shEZH2 and GSK126 treatment, respectively, and the expression of 20 genes was increased or decreased by both shEZH2 and GSK126 treatment. CXCL9 and CXCL10 were at the top 1 and 3 positions among the increased genes in the arrays (Fig. 2h). Altogether, the data indicate that H3K27me3 specific methyltransferase and demethylase preferentially and predominantly regulate T<sub>H</sub>1-type chemokine repression in primary ovarian cancer cells.

DNA methylation regulates gene expression through DNA methyltransferases (DNMTs). We treated two primary and one established ovarian cancer cell line with 5-AZA-dC. 5-AZA-dC treatment increased CXCL9 and CXCL10 mRNA and protein expression (Fig. 3a–c and Extended Data Fig. 4a–d). *IRF1* and *IFNGR2* were not affected by 5-AZA-dC treatment (Extended Data Fig. 4e, f). Specific knockdown of DNMT1 (Extended Data Fig. 4g) increased CXCL9 and CXCL10 mRNA and protein expression (Fig. 3d–f), but had no effect on *IFNGR2* (Extended Data Fig. 4h). To demonstrate DNA methylation status in the *CXCL10* locus, we carried out bisulfite genomic sequencing on the *CXCL10* gene locus and analysed the methylation at the *CXCL10* gene loci containing the STAT1-binding site (Fig. 3g and



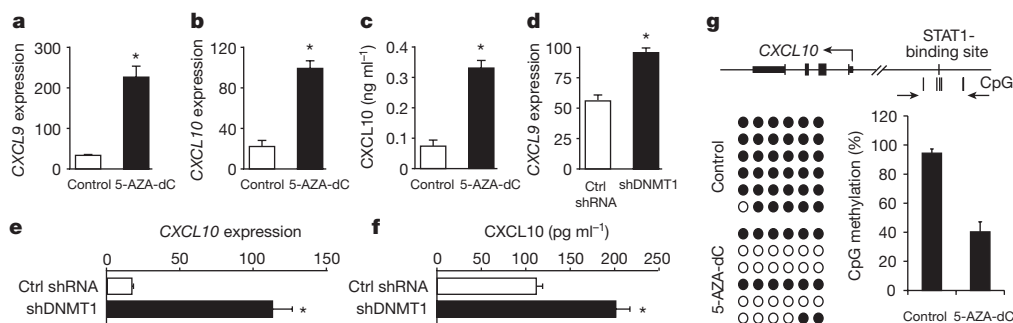
**Figure 2 | Histone methyltransferase represses T<sub>H</sub>1-type chemokines.** **a, b**, Effects of EZH2 knockdown on cancer T<sub>H</sub>1-type chemokine expression. shEZH2- and control-vector-expressing primary ovarian cancer cells were stimulated with IFN $\gamma$ , and chemokines were quantified by real-time PCR (**a**) or ELISA (**b**) (mean  $\pm$  s.d.,  $n = 6$ ,  $*P < 0.05$ , Wilcoxon test). **c**, Effect of shEZH2 on H3K27me3 occupancy at T<sub>H</sub>1-type chemokine promoters. H3K27me3 ChIP was performed in shEZH2-expressing primary ovarian cancer cells. H3K27me3 levels at the gene promoters of CXCL9 and CXCL10 were normalized to the input (mean  $\pm$  s.d.,  $n = 5$ ,  $*P < 0.05$ , Wilcoxon test). **d**, Effects of GSK126 on cancer T<sub>H</sub>1-type chemokine expression. Primary ovarian cancer cells were pre-treated with GSK126 and stimulated with IFN $\gamma$  for 24 h. Chemokine transcripts were quantified by real-time PCR (mean  $\pm$  s.e.m.,  $n = 6$ ,  $*P < 0.05$ , Wilcoxon test). **e–g**, Effects of JMJD3 on cancer T<sub>H</sub>1-type chemokine expression. Primary ovarian cancer

cells were transfected with plasmids encoding HA-JMJD3 (**e**), or with a lentiviral shJMJD-expressing vector (**f**) or pre-treated with GSK-J4 (**g**), and stimulated with IFN $\gamma$  for 24 h. Chemokine transcripts were quantified by real-time PCR (mean  $\pm$  s.d.,  $n = 5$ ,  $*P < 0.05$ , Wilcoxon test). **h**, Effects of EZH2 and H3K27me3 manipulation on global changes of gene expression. Primary ovarian cancer cells were treated with GSK126, or transduced with a shEZH2 lentivirus, and stimulated with IFN $\gamma$ . Global gene expression analysis was done by microarray. Heat map shows differential upregulation and downregulation of genes in GSK126-treated cells and in shEZH2 cells compared to controls. The 20 overlapped differentially expressed genes in both shEZH2- and GSK126-treated cells are shown at the top. Two repeats per condition. Gene expression levels are represented as relative values normalized to GAPDH mRNA level.

Extended Data Fig. 4i). The genomic location of bisulfite sequencing was 5' upstream of the CXCL10 promoter (–5 kb to –4.7 kb). 5-AZA-dC treatment reduced DNA methylation of the CXCL10 gene locus (Fig. 3g). Thus, DNA methylation regulates T<sub>H</sub>1-type chemokine expression.

We examined the relationship between EZH2/H3K27me3 and DNA methylation in the regulation of chemokine expression. We treated primary ovarian cancer cells with 5-AZA-dC or GSK126, and found

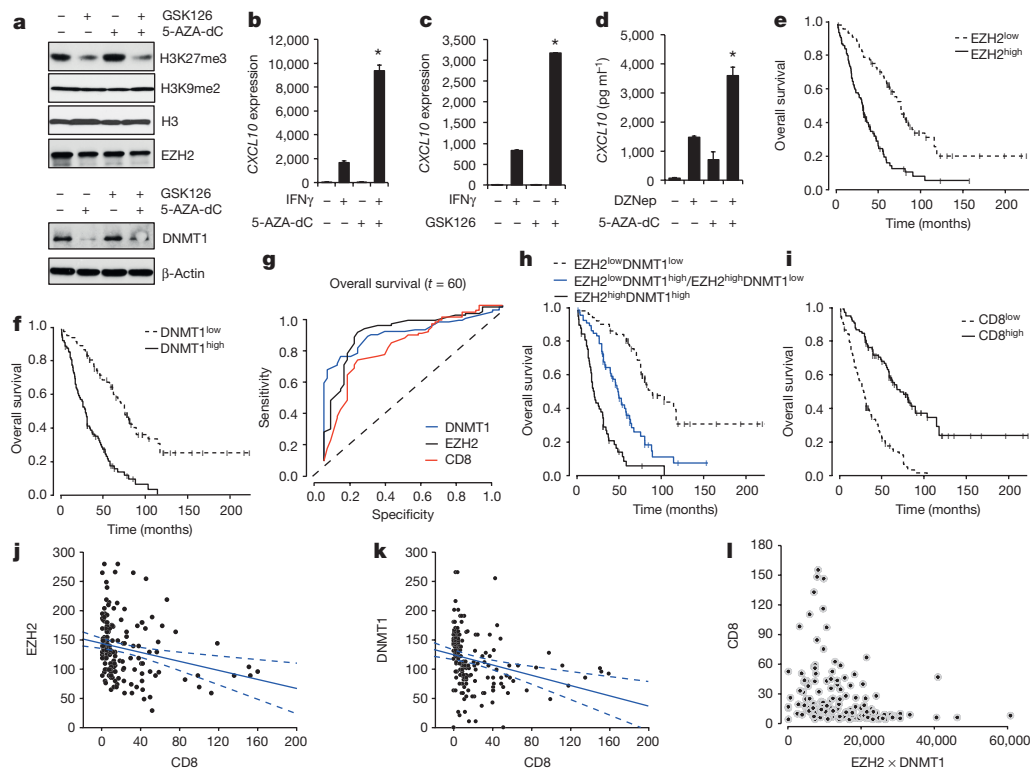
that 5-AZA-dC reduced DNMT1 expression but had no effect on H3K27me3 or H3K9me2, whereas GSK126 reduced H3K27me3 but had no effect on DNMT1 expression (Fig. 4a). Thus, DNA methylation and EZH2/H3K27me3 histone modification may not reciprocally regulate at the protein level in primary ovarian cancer. We explored whether EZH2/H3K27me3 and DNA methylation could independently mediate T<sub>H</sub>1-type chemokine repression. We examined the effects of 5-AZA-dC on shEZH2 cells, and observed that 5-AZA-dC enhanced



**Figure 3 | DNA methylation regulates T<sub>H</sub>1-type chemokine expression.** **a–c**, Effects of 5-AZA-dC on T<sub>H</sub>1-type chemokine expression. Primary ovarian cancer cells were treated with 5-AZA-dC and IFN $\gamma$  for 24 (**a, b**) or 48 (**c**) hours. CXCL9 and CXCL10 transcripts (**a, b**) and protein (**c**) were quantified by real-time PCR or ELISA (mean  $\pm$  s.d.,  $n = 7$ ,  $*P < 0.05$ , Wilcoxon test). **d–f**, Effects of shDNMT1 on T<sub>H</sub>1-type chemokine expression. Primary ovarian cancer cells were transduced with shDNMT1 or non-target shRNA and stimulated with IFN $\gamma$ . Chemokine transcripts (**d, e**) and protein (**f**) were quantified by real-time PCR and ELISA

(mean  $\pm$  s.e.m.,  $n = 5$ ,  $*P < 0.05$ , Wilcoxon test). **g**, Effects of 5-AZA-dC on DNA methylation on the promoter of CXCL10. Schematic diagram of CXCL10 gene locus is shown (top). Primary ovarian cancer cells were treated with 5-AZA-dC. DNA methylation at CpG sites was quantified by bisulfite sequencing. Results are shown as the percentage of methylation. Filled circle, methylated; open circle, unmethylated. The arrows indicate the locations of primers. All gene expressions are represented as relative values normalized to GAPDH mRNA level.





**Figure 4 | EZH2/H3K27 and DNMT1 interaction and its impact on clinical outcome.** **a**, Effects of GSK126 and 5-AZA-dC on H3K27me3 and DNMT1. Primary ovarian cancer cells were treated with GSK126 and/or 5-AZA-dC for 48 h. Histone marks and DNMT1 were detected by western blotting. **b–d**, Effects of 5-AZA-dC and GSK126 on tumour CXCL10. shEZH2 (**b**), shDNMT1 (**c**) expressing cells or primary ovarian cancer cells (**d**) were treated with 5-AZA-dC, GSK126 and/or DZNep for 1–2 days followed by IFN $\gamma$  stimulation for an additional 24 h (**b**, **c**) or 48 h (**d**). CXCL10 was quantified by real-time PCR (**b**, **c**) and ELISA (**d**) (mean  $\pm$  s.e.m.,  $n = 5–6$ , \* $P < 0.05$ , Wilcoxon test). **e**, **f**, The association between EZH2 (**e**), DNMT1 (**f**) and overall survival;  $P < 0.0001$ . **g**, Impact of EZH2, DNMT1 and CD8 $^{+}$  T cells on overall survival. The ROC curve

analysis was applied to evaluate the predictive accuracy of each marker for overall survival ( $t = 60$  months). **h**, Impact of EZH2 and DNMT1 on overall survival. EZH2 $^{low}$ DNMT1 $^{low}$  ( $n = 50$ ), EZH2 $^{high}$ DNMT1 $^{high}$  ( $n = 57$ ), EZH2 $^{low}$ DNMT1 $^{high}$  ( $n = 32$ ), and EZH2 $^{high}$ DNMT1 $^{low}$  ( $n = 66$ ) were compared. Log-rank test,  $P < 0.0001$ , EZH2 $^{low}$ DNMT1 $^{low}$  versus EZH2 $^{high}$ DNMT1 $^{high}$ . **i**, The relationship between intratumoural CD8 $^{+}$  T cells and overall survival;  $P < 0.0001$ . **j–l**, The Pearson correlation between intratumoural CD8 $^{+}$  T cells and EZH2 ( $n = 169$ ,  $r = -0.24$ ,  $P = 0.002$ ) (**j**) and DNMT1 ( $n = 170$ ,  $r = -0.28$ ,  $P = 0.0003$ ) (**k**) and the accumulating levels of EZH2 and DNMT1 ( $n = 168$ ,  $r = -0.32$ ,  $P < 0.0001$ ) (**l**). Gene expression levels are represented as relative values normalized to GAPDH mRNA level.

CXCL10 expression (Fig. 4b). Similarly, GSK126 increased CXCL10 expression in shDNMT1 cells (Fig. 4c). Primary ovarian cancer cells released higher levels of CXCL10 in the combinatorial treatment of DZNep and 5-AZA-dC compared to either treatment alone (Fig. 4d). These data suggest that H3K27me3 and DNA methylation can independently repress tumour T $_{H1}$ -type chemokine expression.

We studied the interaction between EZH2 and DNMT1 in the ovarian cancer microenvironment and its clinical relevance (Extended Data Table 1). We quantified the nuclear levels of EZH2 and DNMT1 (Extended Data Fig. 5a, b) via immunohistochemistry in human ovarian cancer tissues by the H-score method<sup>11</sup> and analysed their impact on patient survival (Extended Data Table 1). Based on the median values of EZH2 and DNMT1 intensity, we divided patients into 'low' and 'high' groups. Overall survival and disease free-interval (DFI) (Extended Data Table 1) were shorter in patients with high levels of EZH2 (Fig. 4e and Extended Data Fig. 5c) and DNMT1 (Fig. 4f and Extended Data Fig. 5d), compared to patients with low levels of these markers. After adjusting for the prognostic clinical factors, overall survival and DFI remained shorter in patients with high EZH2 and DNMT1 expression (Extended Data Table 2). Tumour EZH2 and DNMT1 were positive predictors of risk of death in univariate (Extended Data Table 1) and multivariate analyses (Extended Data Table 2). The reduced overall survival and DFI were more pronounced in the combination of high levels of EZH2 and DNMT1 than EZH2 or DNMT1 alone (Extended Data Table 3).

To estimate the performance of EZH2 and DNMT1 on predicting survival, we used the time-dependent receiver operating characteristic

(ROC) curve analysis<sup>20</sup>. The area under the ROC curve (AUC) is calculated to evaluate the predictive accuracy of each marker for estimating survival. The analysis revealed similar AUCs for EZH2 and DNMT1 in predicting overall survival and DFI (Fig. 4g, Extended Data Fig. 5e and Extended Data Table 4). The expression of EZH2 and DNMT1 may be equally important in ovarian cancer pathology. Next, we evaluated significance of the two parameters for ovarian cancer survival. Given the negative and independent impact of EZH2 and DNMT1, we reasoned that both parameters are more efficient at predicting survival than individually. Indeed, EZH2 $^{high}$ DNMT1 $^{high}$  patients experienced a shorter overall survival and DFI (Extended Data Table 3) than EZH2 $^{low}$ DNMT1 $^{low}$  patients in univariate and multivariate analysis (Fig. 4h). Patients with mixed patterns had similar and moderate survival (Fig. 4h and Extended Data Fig. 5f). Thus, the combination of EZH2 and DNMT1 allows for improved prognostic stratification of ovarian cancer survival as compared to EZH2 or DNMT1 alone. Next, we quantified intratumoural CD8 $^{+}$  T cells by immunohistochemistry (Extended Data Fig. 5g), and found that there was a correlation between tumour CD8 $^{+}$  T-cell content and survival (Extended Data Tables 1 and 2). We divided the patients into two groups based on the median values of CD8 $^{+}$  T-cell number. Patients with more CD8 $^{+}$  T cells experienced longer overall survival and DFI scores compared to those with fewer CD8 $^{+}$  T cells<sup>21,22</sup> (Fig. 4i and Extended Data Fig. 5h). ROC curve analysis revealed similar power of CD8 $^{+}$  T cells in predicting overall survival (Fig. 4g) and DFI (Extended Data Fig. 5e).

Finally, we observed that the levels of EZH2 (Fig. 4j) and DNMT1 (Fig. 4k) were inversely associated with CD8 $^{+}$  T cells. The accumulating



levels of EZH2 and DNMT1 negatively correlated with the number of intratumoural CD8<sup>+</sup> T cells (Fig. 4l). Altogether, the results suggest that tumours co-opt certain epigenetic pathways, silence T<sub>H</sub>1-type chemokine expression and repress T-cell tumour homing, which functions as a novel immune-evasion mechanism.

To maintain cellular identity, through epigenetic mechanisms, key 'stemness' genes may be repressed in somatic cells, and key effector genes may be silenced in stem cells<sup>23,24</sup>. Tumour cells may gain stem cell properties<sup>11</sup>. This may explain why T<sub>H</sub>1-type chemokines, as effector genes, are repressed in cancer. Epigenetic silencing is an intrinsic tumorigenic mechanism<sup>1</sup>. We propose a unifying model of cancer in which epigenetic dysregulation has dual biological and immunological roles in supporting tumour progression. Cancer immunotherapies<sup>4,5,7,8,25</sup> and classic therapies<sup>26–28</sup> rely on efficient T-cell tumour trafficking, and may induce and/or expand mutated TAA-specific T cells<sup>29,30</sup>. Applicably, our work suggests that epigenetic reprogramming may condition tumours from poor T-cell infiltration to rich T-cell infiltration<sup>6</sup>, and ultimately potentiate cancer therapy (Extended Data Fig. 5i).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 9 December 2014; accepted 24 August 2015.**

**Published online 26 October; corrected online 11 November 2015**

(see full-text HTML version for details).

1. Timp, W. & Feinberg, A. P. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nature Rev. Cancer* **13**, 497–510 (2013).
2. Curiel, T. J. *et al.* Blockade of B7–H1 improves myeloid dendritic cell-mediated antitumor immunity. *Nature Med.* **9**, 562–567 (2003).
3. Zou, W. & Chen, L. Inhibitory B7-family molecules in the tumour microenvironment. *Nature Rev. Immunol.* **8**, 467–477 (2008).
4. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nature Rev. Cancer* **12**, 252–264 (2012).
5. Rosenberg, S. A., Restifo, N. P., Yang, J. C., Morgan, R. A. & Dudley, M. E. Adoptive cell transfer: a clinical path to effective cancer immunotherapy. *Nature Rev. Cancer* **8**, 299–308 (2008).
6. Spranger, S., Bao, R. & Gajewski, T. F. Melanoma-intrinsic  $\beta$ -catenin signalling prevents anti-tumour immunity. *Nature* **523**, 231–235 (2015).
7. Scholler, J. *et al.* Decade-long safety and function of retroviral-modified chimeric antigen receptor T cells. *Sci. Transl. Med.* **4**, 132–153 (2012).
8. Brentjens, R. J. *et al.* CD19-targeted T cells rapidly induce molecular remissions in adults with chemotherapy-refractory acute lymphoblastic leukemia. *Sci. Transl. Med.* **5**, 177ra38 (2013).
9. Tan, J. *et al.* Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev.* **21**, 1050–1063 (2007).
10. Knutson, S. K. *et al.* Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. *Mol. Cancer Ther.* **13**, 842–854 (2014).
11. Cui, T. X. *et al.* Myeloid-derived suppressor cells enhance stemness of cancer cells by inducing microRNA101 and suppressing the corepressor CtBP2. *Immunity* **39**, 611–621 (2013).
12. Curiel, T. J. *et al.* Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nature Med.* **10**, 942–949 (2004).
13. Kryczek, I. *et al.* B7–H4 expression identifies a novel suppressive macrophage population in human ovarian carcinoma. *J. Exp. Med.* **203**, 871–881 (2006).
14. Kryczek, I. *et al.* Human TH17 cells are long-lived effector memory cells. *Sci. Transl. Med.* **3**, 104ra100 (2011).
15. McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108–112 (2012).
16. Zhang, Y. *et al.* The polycomb repressive complex 2 governs life and death of peripheral T cells. *Blood* **124**, 737–749 (2014).
17. Cao, R. & Zhang, Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr. Opin. Genet. Dev.* **14**, 155–164 (2004).
18. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731–734 (2007).
19. Kruidenier, L. *et al.* A selective jumoni H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404–408 (2012).
20. Heagerty, P. J., Lumley, T. & Pepe, M. S. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**, 337–344 (2000).
21. Zhang, L. *et al.* Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N. Engl. J. Med.* **348**, 203–213 (2003).
22. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960–1964 (2006).
23. Doege, C. A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652–655 (2012).
24. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488**, 409–413 (2012).
25. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
26. Zitvogel, L., Kepp, O. & Kroemer, G. Decoding cell death signals in inflammation and immunity. *Cell* **140**, 798–804 (2010).
27. Park, S. *et al.* The therapeutic effect of anti-HER2/neu antibody depends on both innate and adaptive immunity. *Cancer Cell* **18**, 160–170 (2010).
28. Twyman-Saint Victor, C. *et al.* Radiation and dual checkpoint blockade activate non-redundant immune mechanisms in cancer. *Nature* **520**, 373–377 (2015).
29. Tran, E. *et al.* Cancer immunotherapy based on mutation-specific CD4<sup>+</sup> T cells in a patient with epithelial cancer. *Science* **344**, 641–645 (2014).
30. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).

**Acknowledgements** This work is supported (in part) by the NIH grants (CA190176; CA123088; CA099985; CA156685; CA193136; CA152470; CA171306; 5P30CA46592), the Rivkin Ovarian Cancer Center, and the Ovarian Cancer Research Fund. We appreciate discussion with B. Richardson. We thank L. Cabrera, D. Postiff, M. Vinco, R. Craig and J. Barikdar for their technical assistance. We are grateful for the professional help from W. Wu and C. Johnson in our Microarray Core and Bioinformatics Core. We particularly appreciate the support of B. Leclair and D. Leclair.

**Author Contributions** D.P. and W.Z. initiated and designed the research. D.P., I.K. and W.Z. wrote the manuscript. D.P., I.K., N.N., S.W., E.Z., L.V., W.W. and W.S. performed experiments. L.Z., I.K. and D.P. analysed data. J.K., R.T., Y.D., S.H.-A., A.M., K.C. and R.L. provided intellectual and technical support, clinical specimens, and clinical and pathological information.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.Z. ([wzou@med.umich.edu](mailto:wzou@med.umich.edu)).

## METHODS

**Ovarian cancer patients, cancer tissue samples and primary ovarian cancer cells.** Patients diagnosed with high-grade serous ovarian carcinomas were recruited for this study. Human subject use in this study was approved by the local Institutional Review Boards and informed consent was obtained from the patients. We collected 186 formalin-fixed, paraffin-embedded ovarian tumour tissue blocks (Extended Data Table 1) and 20 fresh ovarian cancer tissues for this study as described previously<sup>2,12,31</sup>. After pathological review, a tissue microarray (TMA)<sup>11,32</sup> was constructed from the most representative area of paraffin-embedded ovarian tumour tissue. For each tumour, a minimum of two representative tumour areas were selected from a haematoxylin and eosin-stained section of a donor block. Core cylinders (1 mm) were punched from each of these areas and deposited into a recipient paraffin block. Consecutive 6-µm-thick TMA sections were cut and placed on charged poly-L-lysine-coated slides for immunohistochemistry analyses. Epithelial cell adhesion molecule-expressing (EpCAM<sup>+</sup>) primary tumour cells, CD3<sup>+</sup>CD45<sup>+</sup> T cells, and CD45<sup>+</sup> immune cells were isolated and sorted from fresh ovarian cancer tissues for functional studies. Primary human ovarian cancer cell lines (OC8 and OC17) were generated from fresh ovarian cancer tissues and/or ascites fluid in our laboratory as previously described<sup>11,13,33</sup>. Three commercialized ovarian cancer cell lines A2780, CAOV3 and ES-2 were included in the study. The cell lines were routinely tested for mycoplasma contamination. All the commercialized cell lines were authenticated by the supplier and used within ten passages. Primary ovarian cancer cells were authenticated in the laboratory.

**Plasmids, shRNA and antibodies.** pCMV-HA-JMJD3 plasmid was obtained from Addgene (#24167). Lentiviral shRNAs (Extended Data Table 5) were provided by the Vector Core at the University of Michigan or provided by A. Chinnaiyan (University of Michigan). Antibodies including monoclonal anti-EZH2 (1:2000, BD Biosciences, 612667), anti-H3K27me3 (1:1000, Millipore, 07-449), H3K9me2 (1:2000, ab1220, Abcam), H3K9me3 (1:500, ab8898, Abcam), H3K4me1 (ab8895, Abcam), H3K4me2 (ab194678, Abcam), H3K4me3 (ab1012, Abcam), anti-Histone H3 (1:2000, Cell Signaling, 9715), anti-HA (1:200, Santa Cruz Biotechnology, sc-805), anti-DNMT1 (1:250, Abcam, ab13537) were used for western blotting. Anti-human CXCR3 (1C6) blocking antibody was prepared from mouse hybridoma (ATCC, #HB-12330) by Hybridoma Core at the University of Michigan.

**Quantitative Real-time RT-PCR and microarray.** RNA extracted with TRIzol (Invitrogen) was used for cDNA synthesis with high capacity cDNA reverse transcript kit (Applied Biosystems, 4374966). Real-time RT-PCR was performed on the Eppendorf Realplex real-time PCR system or StepOne Plus real-time PCR system (Life Technologies). The mRNA was quantified and normalized to *GAPDH*. The specific primers used for Real-time PCR are listed in the Extended Data Table 5.

Gene expression microarray was carried out at the University of Michigan Microarray Core Facility using Affymetrix Human Gene ST 2.1 Chip according to the standard protocol. Data was analysed by the University of Michigan Bioinformatics Core Facility. Two biological replicates of each sample were prepared independently from primary human ovarian cancer cells for gene expression profiling. Genes with >1.5-fold changes were selected for further analysis.

**Chromatin immunoprecipitation (ChIP).** ChIP assay was performed as described previously<sup>11,32</sup>. Sonication was performed with the Misonix 4000 water bath sonication unit at 15% amplitude for 10 min. ChIP-enriched chromatin was analysed by real-time PCR with SYBR Green Master Mix, and normalized to the input. The ChIP primers were listed in the Extended Data Table 5.

**Immunofluorescence staining (IF) and immunohistochemistry staining (IHC).** Mouse tumour xenografts were harvested and either fixed in 10% buffered formalin and embedded in paraffin for immunohistochemistry assay, or immediately frozen and embedded in OCT compound embedding medium for immunofluorescence staining. The immunofluorescence staining was performed with the monoclonal anti-CD8 (HIT8a, #550372, BD Biosciences). Alexa Fluor 488 or Alexa Fluor 568 conjugated anti-mouse secondary antibodies (Life Technologies) were used. Staining with isotype antibody was used as negative control. T cells were counted manually at 10–20 high-power fields under a fluorescence microscope (Leica). Immunohistochemical staining on human TMA slides or frozen ovarian cancer tissue sections was performed on a DAKO autostainer (DAKO) using DAKO LSAB<sup>+</sup> and diaminobenzidine (DAB) as the chromogen. Serial sections of de-paraffinized TMA sections were labelled with rabbit anti-human EZH2 antibody (Clone ZMD.309, 18-7395, 1:200, Life Technologies), rabbit anti-DNMT1 (Clone H-300, sc-20701, 1:50, Santa Cruz Biotechnology), and mouse anti-human CD8 antibody (DAKO, M7103). Cores from several normal organ tissues were used as tissue controls on each slide.

The cores were quantified and analysed for the expression of EZH2 and DNMT1 with an Aperio imaging system (Genetix) and for the numbers of intratumoral

CD8<sup>+</sup> T cells in 0.1 mm<sup>2</sup> tumour. The specimens were digitalized with an automated platform (Aperio Technologies) and ScanScope XT and Spectrum Plus using TMA software version 9.1 scanning system. Cores were scored in high resolution at 40×. Tissue sections were scored manually on a computer screen, and a mean score for duplicate cores from each individual was calculated. Any discrepancies were resolved by subsequent consultation with diagnostic pathologists. Epigenetic marks, EZH2 and DNMT1 were localized in the nuclei, and were scored using the H-score method<sup>11,32</sup>. The tissues were divided into high and low epigenetic mark expression based on the median value of the expression level per core. The H-score is a method of assessing the extent of nuclear immunoreactivity. The H-score takes into account the percentage of positive cells (0–100%) in each intensity category (0–3+) and computes a final score, on a continuous scale between 0 and 300. The score is obtained by the formula: 3 × percentage of strongly staining nuclei + 2 × percentage of moderately staining nuclei + percentage of weakly staining nuclei; giving a range of 0 to 300<sup>11</sup>. Staining with isotype antibody was used as negative control.

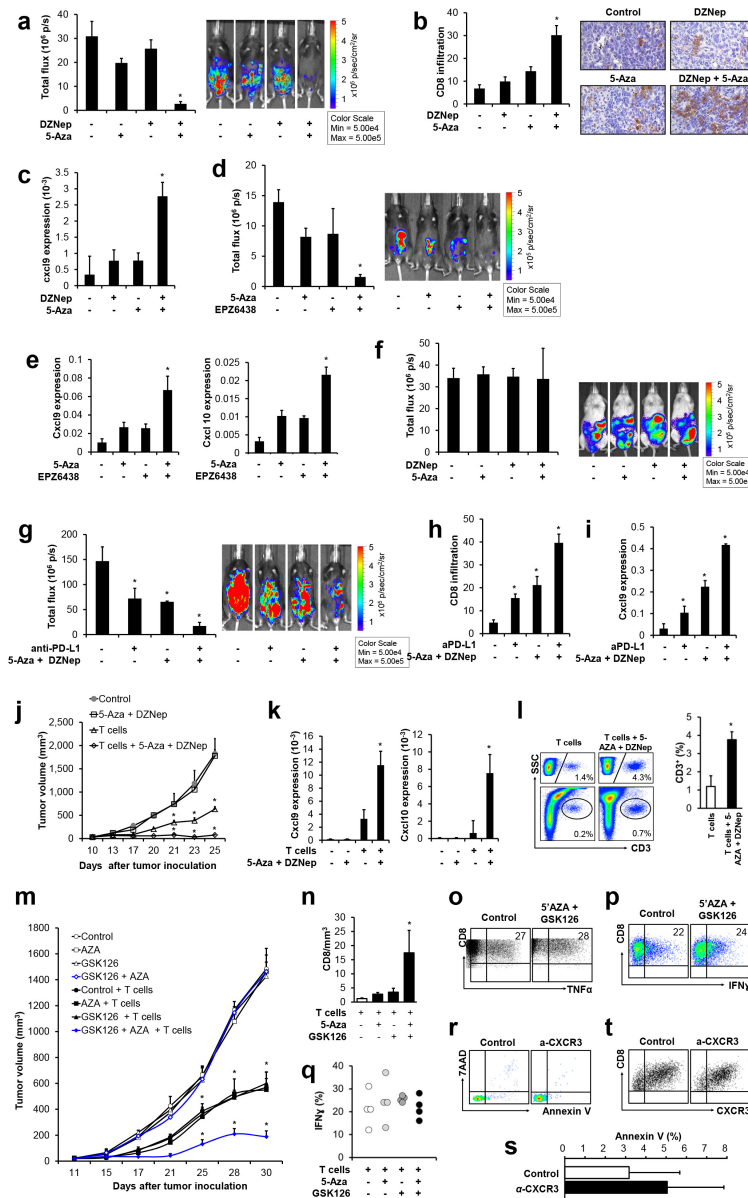
**In vivo tumour progression and immunotherapy models.** ID8 mouse ovarian cancer cells have been described previously<sup>34</sup>. ID8 cells (5 × 10<sup>5</sup>) were injected into peritoneal cavity of NSG mice or C57BL/6 mice (6–8 weeks old, Jackson Lab)<sup>11,32</sup>. Tumour progression was monitored 2–3 times per week by Xenogen IVIS Spectrum *in vivo* bioluminescence imaging system (PerkinElmer). Tumour volume was calculated based on the total flux (photons per second). Tumour-bearing mice were treated (i.p.) with 5 mg kg<sup>−1</sup> DZNep (SML0305, Sigma), 50 mg kg<sup>−1</sup> EPZ6438 (E-7438, Active Biochem), 0.2 mg kg<sup>−1</sup> 5-AZA-dC (A3656, Sigma), or 10 mg kg<sup>−1</sup> anti-PD-L1 (B7-H1, clone 10F9G2, BE0101, Bio X Cell) three times per week for two weeks. In some cases, tumour was dissected for the analysis of chemokine production or T-cell infiltration as indicated.

In an adoptive T-cell therapeutic model, human TAA-specific CD8<sup>+</sup> T cells were generated *in vitro* and primary human ovarian cancer cells were inoculated subcutaneously into the flanks of NSG mice<sup>11,32</sup>. TAA-specific CD8<sup>+</sup> T cells (7 × 10<sup>6</sup>) were intravenously transfused into tumour-bearing mice. DZNep (5 mg kg<sup>−1</sup>), GSK126 (30 mg kg<sup>−1</sup>), and 5-AZA-dC (0.2 mg kg<sup>−1</sup>) treatments were started before T-cell transfusion by intraperitoneal administration three times per week. In some cases, mice received CD8<sup>+</sup> T cells which were pre-incubated with anti-CXCR3 for 1 h before *in vivo* transfusion, followed by intraperitoneal administration of 500 µg anti-CXCR3 three times per week. Tumour growth was monitored and recorded. Tumour cells and tumour infiltrating immune cells were isolated and studied by FACS, real-time PCR and/or immunohistochemistry. All animal protocols were approved by the University of Michigan Committee on Use and Care of Animals (UCUCA).

**Statistical analysis.** Wilcoxon rank-sum tests were used to compare two independent groups, and for paired groups, Wilcoxon signed rank tests were used for the comparison. Correlation coefficients (Spearman correlation, denoted by  $\rho$ , for ordinal data and Pearson correlation, denoted by  $r$ , for continuous data), together with a  $P$  value (null hypothesis is that  $\rho/r$  is zero), were computed to measure the degree of association between biomarkers. log-rank test was used to compare time to tumour initiation between two groups. Overall patient survival was defined as the time from date of diagnosis to disease-related death. Survival functions were estimated by Kaplan–Meier methods.

Cox's proportional hazards regression was performed to model survival as a function of EZH2, DNMT1 and CD8<sup>+</sup> T cells. The data were analysed as continuous or categorized values and classified as low and high based on the median values, or the combination of EZH2 and DNMT1 (classified as EZH2<sup>high</sup>DNMT1<sup>high</sup> and EZH2<sup>low</sup>DNMT1<sup>low</sup>), after adjusting for age and stage. We assessed the adequacy of the Cox regression model. Graphical and numerical methods were described<sup>35</sup>. We used ROC analysis<sup>20</sup> to evaluate the predictive accuracy of the levels of EZH2 and DNMT1, and CD8<sup>+</sup> T cells for 60-month survival. All analyses were done using SAS 9.3 software.  $P < 0.05$  considered as significant. No statistical methods were used to predetermine sample size. Sample size was determined on the basis of animal experimental trials and in consideration of previous publications on similar experiments to allow for confident statistical analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment unless state differently.

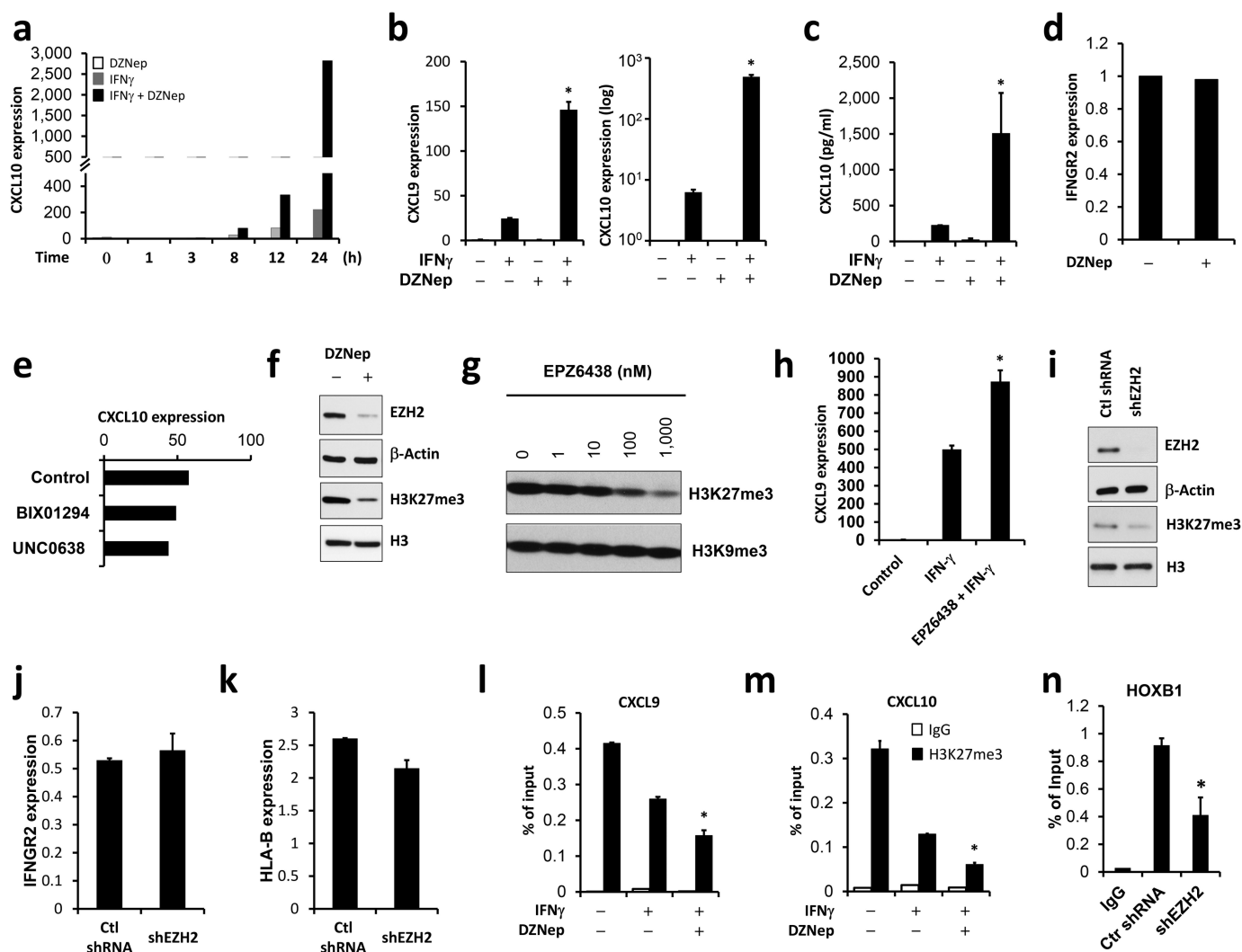
31. Zou, W. *et al.* Stromal-derived factor-1 in human tumors recruits and alters the function of plasmacytoid precursor dendritic cells. *Nature Med.* **7**, 1339–1346 (2001).
32. Kryczek, I. *et al.* IL-22CD4 T cells promote colorectal cancer stemness via STAT3 transcription factor activation and induction of the methyltransferase DOT1L. *Immunity* (2014).
33. Kryczek, I. *et al.* Expression of aldehyde dehydrogenase and CD133 defines ovarian cancer stem cells. *Int. J. Cancer* **130**, 29–39 (2012).
34. Roby, K. F. *et al.* Development of a syngeneic mouse model for events related to ovarian cancer. *Carcinogenesis* **21**, 585–591 (2000).
35. Lin, D. Y., Wei, L. J. & Ying, Z. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572 (1993).



**Extended Data Figure 1 | Epigenetic reprogramming alters immunotherapy.** **a–c**, Effects of DZNep and 5-AZA-dC on ID8 mouse ovarian cancer progression. The ID8 tumour-bearing mice (C57BL/6) were treated with DZNep and 5-AZA-dC. **a**, Tumour growth was recorded by bioluminescence imaging and quantified by calculating the total flux (photons per second). The representative images and tumour volume at day 24 are shown. Day 0, tumour inoculation. **b**, Tumour-infiltrating CD8 $^+$  T cells were quantified by immunohistochemistry staining (IHC) and expressed as the mean  $\pm$  s.e.m. per high-power field. **c**, Tumour *Cxcl9* mRNA was quantified by real-time PCR. Mean  $\pm$  s.e.m.;  $n = 5$  per group;  $*P < 0.05$ , Mann–Whitney *U*-test as compared to the control group or the group treated with DZNep or 5-AZA-dC. **d, e**, Effects of EPZ6438 and 5-AZA-dC on ID8 mouse ovarian cancer progression. The ID8 tumour-bearing mice (C57BL/6) were treated with EPZ6438 and 5-AZA-dC. **d**, Tumour growth was recorded by bioluminescence imaging and quantified by calculating the total flux (photons per second). The representative images and tumour volume at day 22 are shown. Day 0, tumour inoculation. Mean  $\pm$  s.e.m.;  $*P < 0.05$ ; one-way ANOVA. **e**, Tumour *Cxcl9* mRNA was quantified by real-time PCR. Mean  $\pm$  s.e.m.;  $*P < 0.05$ ; one-way ANOVA. **f**, Effects of DZNep and 5-AZA-dC on ovarian tumour progression in NSG mice. Mouse ID8-luc ovarian cancer cells were intraperitoneally inoculated into NSG mice. Tumour-bearing mice were treated with DZNep and 5-AZA dC. Tumour growth was recorded and quantified by bioluminescence imaging. The representative bioluminescence images on day 15 (days after tumour inoculation) are shown.  $n = 4$  per group; mean  $\pm$  s.e.m. **g–i**, Effects of DZNep and 5-AZA-dC on anti-PD-L1 immunotherapy. Anti-PD-L1 (10 mg/kg) was given to ID8 tumour-bearing mice (C57BL/6) with or without DZNep and

5-AZA-dC (5-Aza). **g**, Tumour growth was recorded. The representative images and tumour volume at day 18 are shown. **h**, Tumour-infiltrating CD8 $^+$  T cells were assessed and expressed as the mean  $\pm$  s.e.m. per high-power field. **i**, Tumour *Cxcl9* gene expression was quantified and expressed as the mean values  $\pm$  s.e.m.  $n = 5$  per group;  $*P < 0.05$ ; Mann–Whitney *U*-test). **j–l**, Effects of DZNep and 5-AZA-dC on T-cell immunotherapy. Autologous tumour-specific human CD8 $^+$  T cells were transfused into ovarian-cancer-bearing NSG mice with or without DZNep and 5-AZA-dC treatment. The mice were treated with anti-CXCR3. **j**, Tumour volume was monitored. **k**, TH1-type chemokine expression was quantified by real-time PCR. **l**, Tumour-infiltrating T cells were determined by FACS. Mean  $\pm$  s.e.m.;  $n = 5$  per group;  $*P < 0.05$  Mann–Whitney *U*-test. **m–q**, Effects of GSK126 and 5-AZA-dC on T-cell immunotherapy. Ovarian-cancer-bearing NSG mice were treated with or without GSK126 and 5-AZA-dC, and received autologous human ovarian-cancer-specific CD8 $^+$  T-cell transfection. Tumour volume (**m**), tumour-infiltrating T cells (**n**) and T-cell cytokine profile (**o–q**) are shown Mean  $\pm$  s.e.m.;  $n = 5$  per group;  $*P < 0.05$ ; Mann–Whitney *U*-test. Total tumour-infiltrating CD8 $^+$  T cells were normalized to the tumour volume (absolute number of CD8 $^+$  T cells per  $\text{mm}^3$  of the tumour). T-cell cytokine profile was determined by FACS via gating on human CD45 $^+$ CD8 $^+$  cells in tumour tissues. The percentages of TNF $\alpha$  $^+$  and IFN $\gamma$  $^+$  cells are shown in CD8 $^+$  cells. Circles represent each mouse from the group. One of two experiments is shown. **r–t**, Effects of anti-CXCR3 on T-cell survival. Tumour-specific human CD8 $^+$  T cells were transfused into NSG mice. The mice were treated with anti-CXCR3. Peripheral blood annexin V $^+$ CD8 $^+$  human T cells (mean  $\pm$  s.e.m.;  $n = 5$  per group) (**r, s**) and spleen CXCR3 $^+$ CD8 $^+$  human T cells (**t**) were determined by FACS.

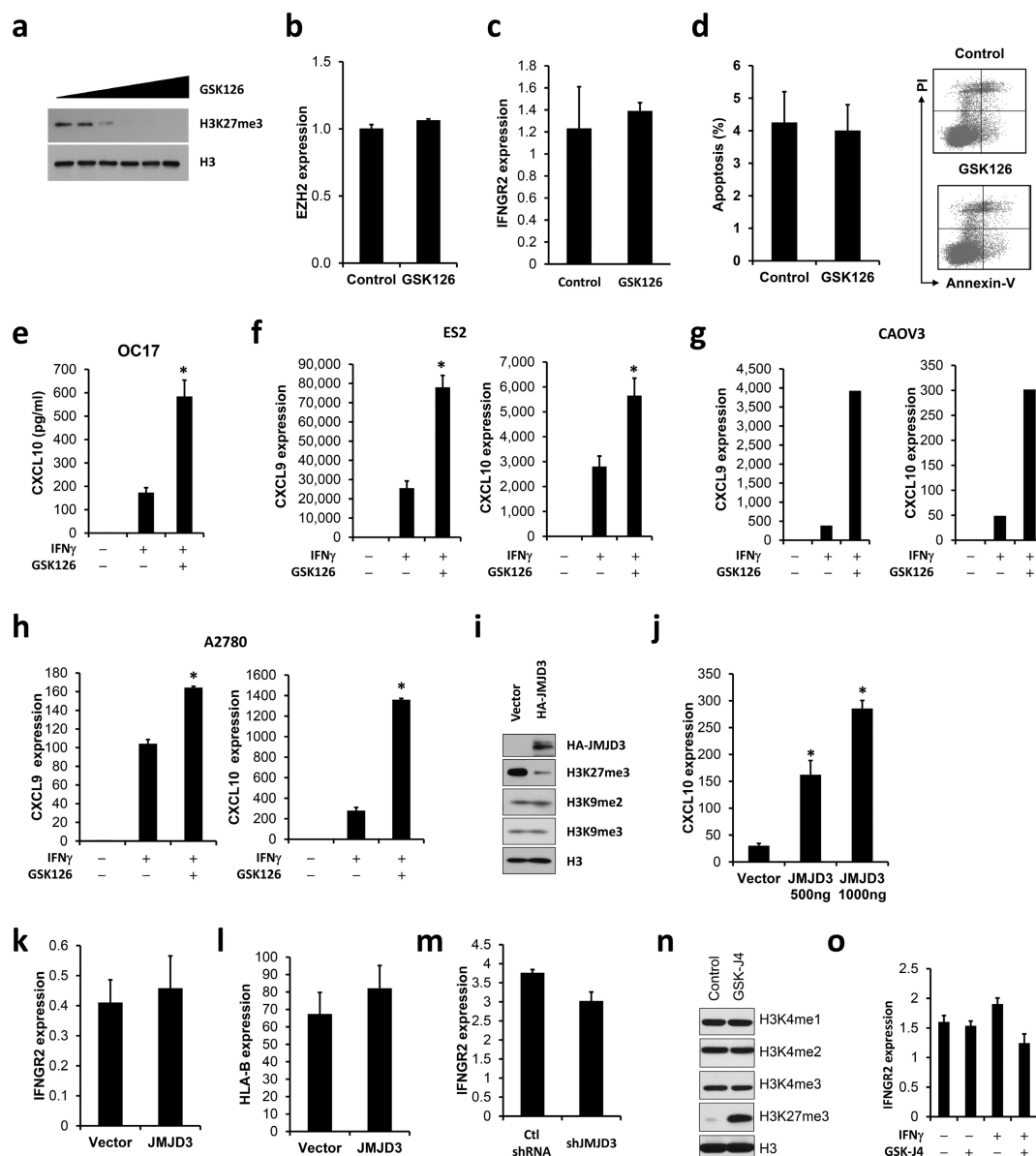




**Extended Data Figure 2 | EZH2 controls T<sub>H</sub>1-type chemokines.** **a**, Effect of DZNep on CXCL10 transcript expression. Primary ovarian cancer cells were treated with DZNep in the presence of IFN $\gamma$  for different time (hours). CXCL10 gene expression was quantified by real-time PCR. One of three experiments is shown. **b**, **c**, Effects of DZNep on ovarian cancer T<sub>H</sub>1-type chemokine expression. Human primary ovarian cancer cells were treated for 24 (**b**) or 48 (**c**) hours with DZNep in the presence of IFN $\gamma$ . CXCL9 and CXCL10 expression was quantified by real-time PCR (**b**) or detected by ELISA (**c**). (mean  $\pm$  s.e.m.;  $n=5$ ,  $*P<0.05$ , Wilcoxon test). **d**, Effects of DZNep on IFNGR2 transcript expression. Primary ovarian cancer cells were treated with DZNep in the presence of IFN $\gamma$  for 24 h. IFNGR2 expression was quantified by real-time PCR. One of three experiments is shown. **e**, Effects of histone methyltransferase G9a/GLP inhibitors on T<sub>H</sub>1-type chemokine expression. Primary ovarian cancer cells were treated with BIX01294 or UNC0638 in the presence of IFN $\gamma$  for 24 h. CXCL10 gene expression was quantified by real-time PCR. One of three experiments is shown. **f**, Effects of DZNep on the expression of EZH2 and H3K27me3. Primary ovarian cancer cells were treated with or without DZNep for 24 h. The levels of EZH2 and H3K27me3 were detected by western blotting. One of three experiments is shown. **g**, **h**, Effects of EPZ6438 on histone marks (**g**) and CXCL9 mRNA expression (**h**). Mouse ID8 ovarian cancer cells were treated with EPZ6438

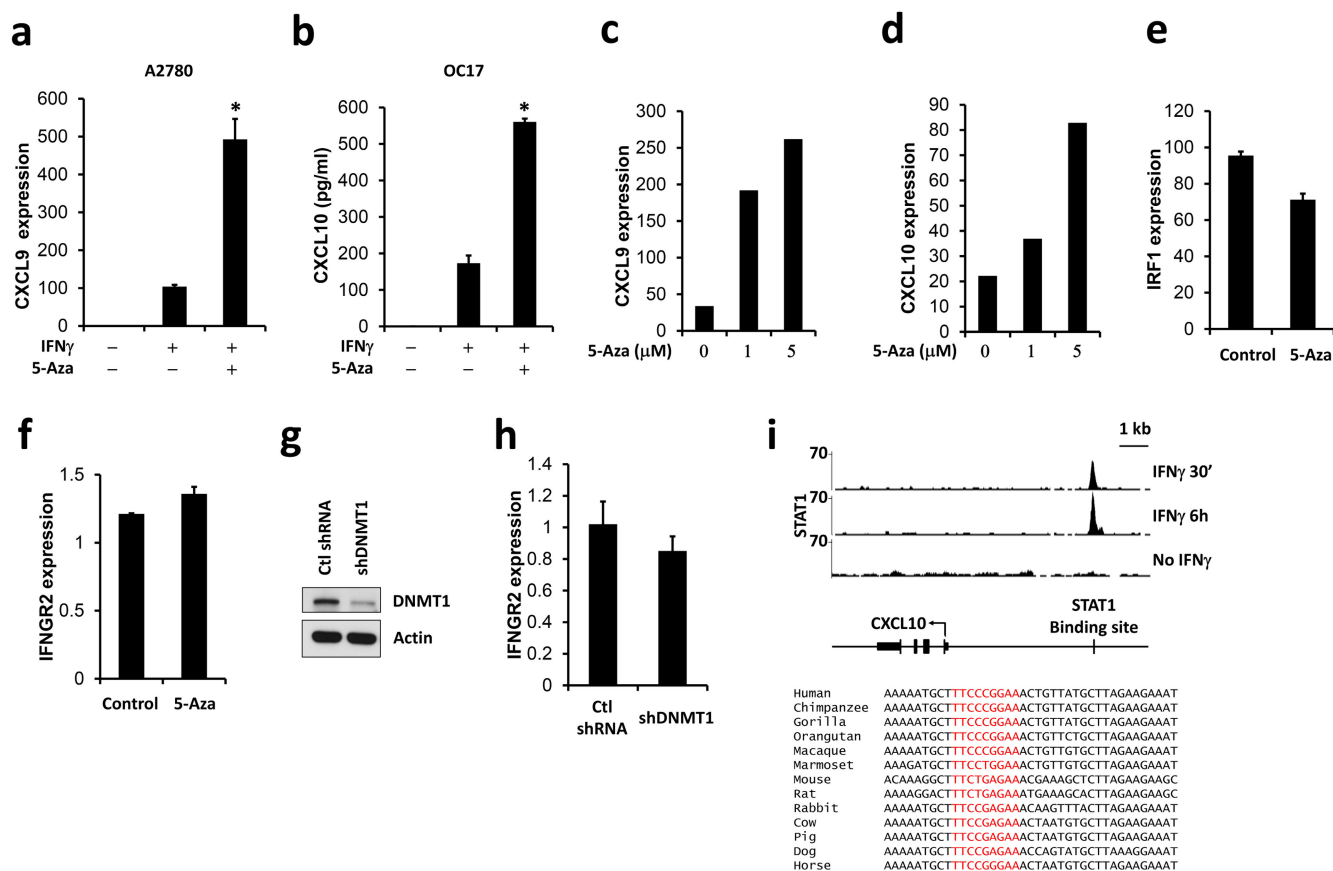
in the presence or absence of IFN $\gamma$  for 48 h. H3K27me3 and H3K9me2 were detected by western blotting. CXCL9 transcripts were quantified by real-time PCR. (Mean  $\pm$  s.e.m.; four repeats,  $*P<0.05$ , Wilcoxon test). **i**, EZH2 knockdown in primary ovarian cancer cells mediated by EZH2 shRNA. Primary ovarian cancer cells were stably transduced with a lentiviral shRNA expressing vector (non-target shRNA (Ctl) or EZH2 shRNA, shEZH2). The levels of EZH2 and H3K27me3 were detected by western blotting. **j**, **k**, Effects of EZH2 knockdown on IFNGR2 and HLA-B gene expression. Primary ovarian cancer cells were stably transduced with non-target shRNA (Ctl) or EZH2 shRNA (shEZH2). IFNGR2 (**j**) and HLA-B (**k**) gene expression was quantified by real-time PCR. (mean  $\pm$  s.e.m.;  $n=4$ ). **l**, **m**, Effect of DZNep on H3K27me3 occupancy at T<sub>H</sub>1-type chemokine promoters. H3K27me3 ChIP assay was performed in primary ovarian cancer cells treated with DZNep with or without IFN $\gamma$ . H3K27me3 levels at the gene promoter of CXCL9 and CXCL10 were normalized to the input (mean  $\pm$  s.e.m.;  $n=5$ ,  $*P<0.05$ , Wilcoxon test). **n**, Effects of EZH2 knockdown on H3K27me3 occupancy at the HOXB1 gene promoter. H3K27me3 ChIP assay was performed in shEZH2- or non-target shRNA-expressing primary ovarian cancer cells. H3K27me3 levels at the gene promoter of HOXB1 were normalized to the input (mean  $\pm$  s.e.m.;  $n=5$ ,  $*P<0.05$ , Wilcoxon test).





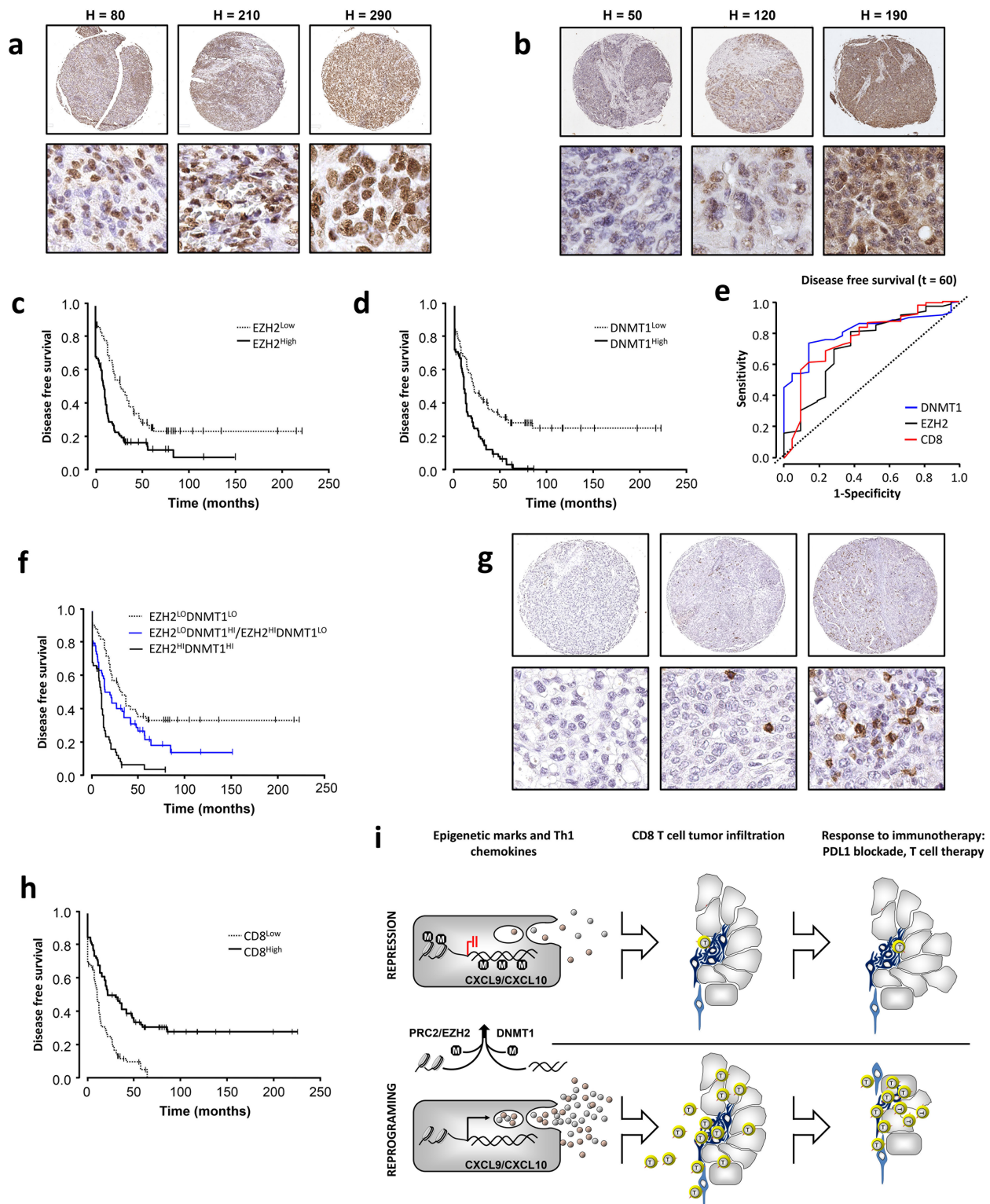
**Extended Data Figure 3 | H3K27-specific methyltransferase and demethylase regulates  $T_H1$ -type chemokine expression.** **a**, Effects of GSK126 treatment on H3K27me3. Primary ovarian cancer cells were treated with GSK126 (0, 0.05, 0.2, 0.5, 2 or 10  $\mu$ M) for 48 h. H3K27me3 was detected by western blotting. One of three experiments is shown. **b**, **c**, Effects of GSK126 on *EZH2* and *IFNGR2* transcript expression. Primary ovarian cancer cells were pre-treated with GSK126 for 48 h and stimulated with IFN $\gamma$  for an additional 24 h. *EZH2* (**b**) and *IFNGR2* (**c**) transcripts were quantified by real-time PCR. Results are expressed as the mean  $\pm$  s.e.m. Results from three experiments with duplicates are shown. **d**, Effects of GSK126 on cell apoptosis. Primary ovarian cancer cells were treated with GSK126 and stained with propidium iodide and annexin V. The percentage of apoptotic cells (annexin V-positive) was quantified. Mean  $\pm$  s.e.m. from three experiments with duplicates. **e-h**, Effects of histone methyltransferase inhibitors on ovarian cancer  $T_H1$ -type chemokine expression. Human primary ovarian cancer cells (OC17) (**e**), or ovarian cancer cell lines (ES-2, CAOV3 and A2780) (**f-h**) were treated with GSK126 with or without IFN $\gamma$ . CXCL9 and CXCL10 expression was measured by ELISA (**e**) or real-time PCR (**f-h**) (mean  $\pm$  s.e.m.,  $n = 5$ ,  $*P < 0.05$ , Wilcoxon test). **i**, Effects of ectopic JMJD3 on histone marks. Primary ovarian cancer cells were transiently transfected with plasmid encoding HA-JMJD3. Histone markers H3K9me2, H3K9me3 and H3K27me3 were detected by western blotting. One of three experiments is shown. **j**, Effect of ectopic expression of JMJD3 on CXCL10 transcript

expression. Primary ovarian cancer cells were transiently transfected with vector or different amounts of HA-JMJD3 plasmids and stimulated with IFN $\gamma$ . CXCL10 transcripts were quantified by real-time PCR. Results are expressed as the mean values  $\pm$  s.e.m. Data represent three independent experiments. ( $n = 5$ ,  $*P < 0.05$ , Wilcoxon test). **k**, **l**, Effects of ectopic JMJD3 on *IFNGR2* (**k**) and *HLA-B* (**l**) gene expression. Primary ovarian cancer cells were transiently transfected with plasmid encoding HA-JMJD3 and stimulated with IFN $\gamma$ . *IFNGR2* and *HLA-B* transcripts were quantified by real-time PCR. Results are expressed as the mean  $\pm$  s.e.m. Data represent three independent experiments. **m**, Effect of JMJD3 knockdown on *IFNGR2* gene expression. Primary ovarian cancer cells were stably transduced with a lentiviral shRNA specific for JMJD3 (shJMJD3) or non-target shRNA (Ctl), and stimulated with IFN $\gamma$ . *IFNGR2* transcripts were quantified by real-time PCR. Results are expressed as the mean  $\pm$  s.e.m. Data represent three independent experiments. **n**, Effects of GSK-J4 on histone marks. Primary ovarian cancer cells were treated with GSK-J4 (10  $\mu$ M) for 48 h. H3K27me3, H3K4me1, H3K4me2 and H3K4me3 were detected by western blotting. One of three experiments is shown. **o**, Effects of GSK-J4 on *IFNGR2* gene expression. Primary ovarian cancer cells were pre-treated with GSK-J4 (10  $\mu$ M) for 48 h and stimulated with IFN $\gamma$  for an additional 24 h. *IFNGR2* transcripts were quantified by real-time PCR. Results are expressed as the mean  $\pm$  s.e.m. Data represent three independent experiments.



**Extended Data Figure 4 | DNA methylation controls  $T_H1$ -type chemokine expression.** **a, b**, Effects of 5-AZA-dC on  $T_H1$ -type chemokine expression. Human ovarian cancer cell line (A2780) or primary ovarian cancer cells (OC17) were treated with 5-AZA-dC and IFN $\gamma$ . CXCL9 and CXCL10 expression were quantified by real-time PCR (**a**) or ELISA (**b**) (mean  $\pm$  s.e.m.,  $n = 6$ ,  $*P < 0.05$ , Wilcoxon test). **c–f**, Effects of 5-AZA-dC on IFN $\gamma$  associated gene expression. Primary ovarian cancer cells (OC8) were treated with 5-AZA-dC and IFN $\gamma$  for 24 h. CXCL9 (**c**), CXCL10 (**d**), IRF1 (**e**) and IFNGR2 (**f**) transcripts were quantified by real-time PCR. Results are expressed as the mean  $\pm$  s.e.m. One of three independent experiments is shown. **g**, DNMT1 knockdown via DNMT1 shRNA. Primary ovarian cancer cells (OC8) were transduced with a lentiviral shRNA specific for DNMT1 (shDNMT1) or non-target shRNA (Ctl). DNMT1 was detected by western blotting. **h**, Effect of DNMT1 knockdown on IFNGR2 expression. Primary ovarian cancer

cells (OC8) were transduced with a lentiviral shRNA specific for DNMT1 (shDNMT1) or non-target shRNA (Ctl) and stimulated with IFN $\gamma$  for 24 h. IFNGR2 transcripts were quantified by real-time PCR. Results are expressed as the mean  $\pm$  s.e.m. Data represent three independent experiments. **i**, STAT1-binding site at CXCL10 gene promoter. STAT1 ChIP-seq data set from ENCODE/SYDH (top panel). K562 cells were treated with IFN $\gamma$  for 30 min or 6 h. Non-treated GM12878 cells were used as control (no IFN $\gamma$ ). Gene Expression Omnibus accession numbers: GSM935487, GSM935488 and GSM935612. STAT1 occupancy at CXCL10 promoter ( $-5,143$  to  $-4,699$  base pairs) is shown as the peaks. Middle panel, schematic diagram of CXCL10 gene locus. STAT1-binding site (TTCCCGGAA) was predicted by TFSEARCH, score = 100. STAT1 ChIP-seq peaks overlap with the predicted STAT1-binding site (indicated as vertical lines). Bottom, homologous STAT1-binding site at CXCL10 gene promoter (Ensemble Genomic alignment).



**Extended Data Figure 5 | EZH2/H3K27 and DNMT1 interaction affects clinical outcome.** **a, b,** Representative images of immunohistochemistry staining of EZH2 (**a**) and DNMT1 (**b**) in human ovarian cancer tissues. The levels of DNMT1 and EZH2 expression in the tumour were assessed by the H-score method. **c, d,** The association between EZH2 (**c**), DNMT1 (**d**) and patient DFS in high-grade serous ovarian cancer. The high and low levels of EZH2 and DNMT1 were determined by the median values (see Extended Data Table 1). **e,** Relative impact of EZH2, DNMT1 and CD8 on patient DFS in high-grade serous ovarian cancer. The time-dependent receiver operating characteristic (ROC) curve analysis was applied to evaluate the predictive accuracy of each marker for DFS. AUC, the area under the ROC curve;  $t = 60$  months. **f,** Impact of the two parameters (EZH2 and DNMT1) on

patient DFS. The analysis was performed on patients with high-grade serous ovarian cancer. Multiple comparisons were performed with log-rank tests.  $EZH2^{low}DNMT1^{low}$  group ( $n = 49$ ) versus  $EZH2^{high}DNMT1^{high}$  ( $n = 55$ )  $P < 0.00001$ . **g,** Representative images of immunohistochemistry staining of CD8 in human ovarian cancer specimen. Intratumoural CD8<sup>+</sup> T cells were shown with anti-CD8<sup>+</sup> staining. The numbers of intratumoural CD8<sup>+</sup> T cells were quantified in high-power fields (40 $\times$ ) (see Methods and Extended Data Table 1). **h,** The relationship between intratumoural CD8<sup>+</sup> T cells and patient DFS in high-grade serous ovarian cancer (see Methods and Extended Data Table 1). **i,** Schematic diagram depicting the relationship among epigenetic T<sub>H</sub>1-type chemokine silencing, effector T-cell trafficking, and tumour immunity, immunotherapy and patient outcome. M, methylation.

Extended Data Table 1 | Patient characteristics and risk factors for overall survival and DFS

		N	P	Hazard Ratio	95% Hazard Ratio Confidence Limits	
Overall survival						
Age: mean (min-max)	59 (19-87)	186	0.004	1.020	1.006	1.033
Stage	I, II	30 (16%)	0.0004	2.955	1.627	5.367
	III, IV	152 (82%)				
	ND	4 (2%)				
EZH2 *	High	103	< 0.0001	3.019	2.083	4.376
	Low	83				
DNMT1 *	High	90	< 0.0001	3.49	2.383	5.11
	Low	83				
CD8 *	High	97	< 0.0001	0.263	0.181	0.384
	Low	89				
Disease free survival						
Age: mean (min-max)	59 (19-87)	178	0.01	1.016	1.004	1.029
Stage	I, II	28 (16%)	0.0004	2.739	1.572	4.772
	III, IV	146 (82%)				
	ND	4 (2%)				
EZH2 *	High	99	< 0.0001	2.112	1.487	2.998
	Low	79				
DNMT1 *	High	86	< 0.0001	2.08	1.46	2.964
	Low	80				
CD8 *	High	90	< 0.0001	0.422	0.298	0.598
	Low	88				

Note: univariate analysis.

\*Categorized values of EZH2, DNMT1 and CD8 are based on the median.



Extended Data Table 2 | Risk factors for overall survival and DFS

<i>Risk factors</i>	<i>P-value</i>	<i>Hazard ratio (HR)</i>	<i>95% HR Confidence Limits</i>	
Overall survival				
EZH2*	< 0.0001	2.657	1.817	3.884
Age	0.0435	1.014	1	1.029
Stage	0.0236	2.038	1.1	3.776
DNMT1*	< 0.0001	3.217	2.182	4.743
Age	0.1096	1.011	0.998	1.025
Stage	0.0038	2.551	1.352	4.814
CD8 *	< 0.0001	0.31	0.211	0.455
Age	0.117	1.011	0.997	1.024
Stage	0.0215	2.134	1.118	4.074
Disease free survival				
EZH2*	0.0004	1.916	1.34	2.738
Age	0.1321	1.01	0.997	1.024
Stage	0.0126	2.096	1.172	3.749
DNMT1*	0.0004	1.929	1.344	2.77
Age	0.0912	1.011	0.998	1.025
Stage	0.0116	2.104	1.181	3.748
CD8*	<.0001	0.471	0.329	0.673
Age	0.1653	1.009	0.996	1.022
Stage	0.0134	2.084	1.164	3.731

Note: multivariate analysis.

\*Categorized values of EZH2, DNMT1 and CD8 are based on the median.

Extended Data Table 3 | The nested multivariate models of overall survival and DFS

<b><i>Overall survival models</i></b>	<b><i>P value</i></b>	<b><i>likelihood rate statistics</i></b>
DNMT1 and EZH2 vs. DNMT1 alone*	< 0.0001	35.5
DNMT1 and EZH2 vs. EZH2 alone*	< 0.0001	47.2
<b><i>Disease free survival</i></b>	<b><i>P value</i></b>	<b><i>likelihood rate statistics</i></b>
DNMT1 and EZH2 vs. DNMT1 alone*	< 0.0001	16.6
DNMT1 and EZH2 vs. EZH2 alone*	< 0.0001	17

\*Categorized values of EZH2, DNMT1 are based on the median.

Extended Data Table 4 | The AUC values of ROCs (t=60) for DNMT1, EZH2 and CD8

	<i><b>DNMT1</b></i> <i><b>AUC ± s.e.</b></i>	<i><b>EZH2</b></i> <i><b>AUC ± s.e</b></i>	<i><b>CD8</b></i> <i><b>AUC ± s.e</b></i>
Overall survival	82.66 ± 3.34	83.63 ± 3.53	75.13 ± 4.18
Disease free survival	81.06 ± 3.91	71.98 ± 6.28	76.53 ± 5.86

Extended Data Table 5 | Sequences for ChIP, PCR primers and lentiviral shRNAs

Primers for quantitative Real-Time PCR		
Gene	Sequence (5'-3')	Sequence (5'-3')
Human CXCL10	CTCCAGTCTCAGCACCATGA	GCTCCCCTCTGGTTTTAAGG
Human CXCL9	GTGGTGTTCTTTTCCTCTTGGG	ACAGCGACCCTTTCTCACTAC
Human IFNGR2	TTCAGTGGCGCCAGTCCCTCA	GGGAGCCTTCTCCTGGGGTCA
human HLA-B	TCCTAGCAGTTGTGGTCATG	TCAAGCTGTGAGAGACACAT
Human EZH2	TTCATGCAACACCCAACACTT	GGTGGGGTCTTTATCCGCTC
Human IRF1	GAGACCCTGGCTAGAGATGC	CATGGCACAGCGAAAGTTGG
Mouse CXCL10	AATGAGGGCCATAGGGAAGC	AGCCATCCACTGGGTAAAGG
Mouse CXCL9	GAGCAGTGTGGAGTTCGAGG	TCCGGATCTAGGCAGGTTTG

Primers for ChIP assay		
Gene	Sequence (5'-3')	Sequence (5'-3')
CXCL10	GGAATGTCTCAGAAAACGTGGGGC	ACCTTCGAGTCTGCAACATGGGA
CXCL9	AGGGTTTCCCCCAGCACAATCA	ACCAGCAAGATGATGCCCAAGAGG
HOXB1	GGGTTGGGAGGGAAGGAAAG	CCCATCCATCTGAGAGCGAC

Lentiviral shRNA		
shRNA	Vector	Source
shEZH2	pGreen-shEZH2-1	Kindly provided by Dr. Arul Chinnaiyan
	pGreen-shEZH2-2	Kindly Provided by Dr. Arul Chinnaiyan
shJMJD3	pGIPz-shJMJD3 (V3LHS_301324)	Vector Core (University of Michigan)
	pGIPz-shJMJD3 (V3LHS_310328)	Vector Core (University of Michigan)
shDNMT1	pGIPZ-DNMT1_V2LHS_113503	Vector Core (University of Michigan)
	pGIPZ-DNMT1_V2LHS_113505	Vector Core (University of Michigan)



# The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions

Elwood A. Mullins<sup>1</sup>, Rongxin Shi<sup>1</sup>, Zachary D. Parsons<sup>1</sup>, Philip K. Yuen<sup>2</sup>, Sheila S. David<sup>2</sup>, Yasuhiro Igarashi<sup>3</sup> & Brandt F. Eichman<sup>1</sup>

Threats to genomic integrity arising from DNA damage are mitigated by DNA glycosylases, which initiate the base excision repair pathway by locating and excising aberrant nucleobases<sup>1,2</sup>. How these enzymes find small modifications within the genome is a current area of intensive research. A hallmark of these and other DNA repair enzymes is their use of base flipping to sequester modified nucleotides from the DNA helix and into an active site pocket<sup>2–5</sup>. Consequently, base flipping is generally regarded as an essential aspect of lesion recognition and a necessary precursor to base excision. Here we present the first, to our knowledge, DNA glycosylase mechanism that does not require base flipping for either binding or catalysis. Using the DNA glycosylase AlkD from *Bacillus cereus*, we crystallographically monitored excision of an alkylpurine substrate as a function of time, and reconstructed the steps along the reaction coordinate through structures representing substrate, intermediate and product complexes. Instead of directly interacting with the damaged nucleobase, AlkD recognizes aberrant base pairs through interactions with the phosphoribose backbone, while the lesion remains stacked in the DNA duplex. Quantum mechanical calculations revealed that these contacts include catalytic charge-dipole and CH– $\pi$  interactions that preferentially stabilize the transition state. We show *in vitro* and *in vivo* how this unique means of recognition and catalysis enables AlkD to repair large adducts formed by yatakemycin, a member of the duocarmycin family of antimicrobial natural products exploited in bacterial warfare and chemotherapeutic trials<sup>6,7</sup>. Bulky adducts of this or any type are not excised by DNA glycosylases that use a traditional base-flipping mechanism<sup>5</sup>. Hence, these findings represent a new model for DNA repair and provide insights into catalysis of base excision.

Alkylation of DNA by endogenous metabolites, environmental toxins and chemotherapeutic agents is a major source of genotoxic damage<sup>8</sup>. By virtue of their positive charge, N3- and N7-alkylpurines are prone to spontaneous depurination at physiological pH, and both N3-methyladenine (3mA) and apurinic/apyrimidinic (AP) sites interfere with DNA replication and transcription<sup>9,10</sup>. As the initial step in the base excision repair pathway, DNA glycosylases remove 3mA and other cationic and neutral nucleobases from the genome. Both enzymatic and non-enzymatic depurination of these lesions have been shown to proceed through a stepwise pathway initiated by cleavage of the glycosidic bond, followed by addition of the nucleophilic water to the oxocarbenium (dR<sup>+</sup>) intermediate<sup>11</sup> (Fig. 1a). The resulting AP site is then converted to an undamaged nucleotide by a common set of lesion-independent base excision repair enzymes<sup>1,2</sup>.

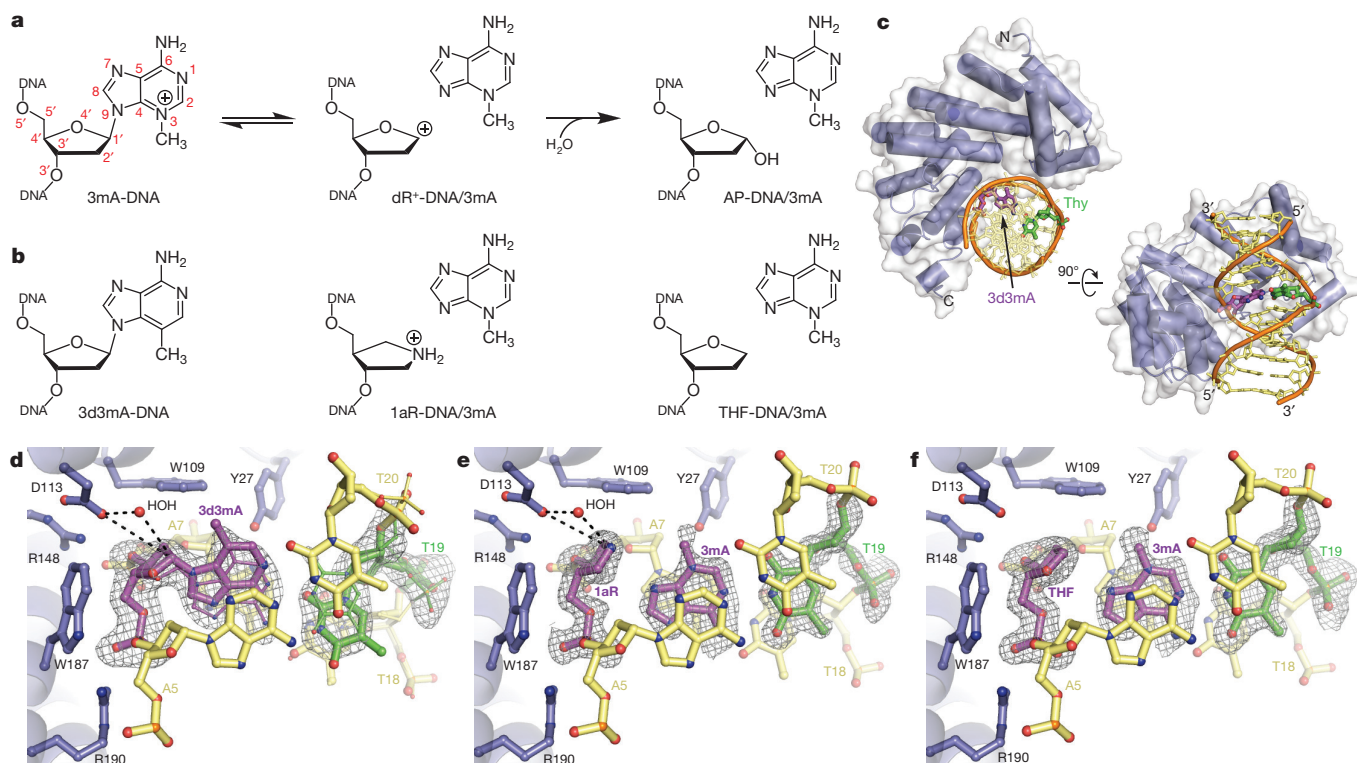
Despite their structural diversity, DNA glycosylases are generally thought to accomplish base flipping through the use of two conserved protein elements—a nucleobase binding pocket and a DNA intercalating residue<sup>5</sup>. The nucleobase binding pocket provides a means of damage recognition through shape and charge complementarity with the modified base, while also creating a scaffold for the residues that catalyse hydrolysis of the N-glycosidic bond. With few exceptions, an aspartate or glutamate residue has a dual role in stabilizing developing

positive charge on the sugar as the glycosidic bond is broken and in deprotonating the nucleophilic water that forms the AP product<sup>12</sup>. Additional residues in the pocket enhance the leaving-group potential of the base. For neutral alkylpurines, this enhancement comes from a general acid that protonates the nucleobase to provide the same instability inherent to cationic purines, which are in essence pre-activated for depurination<sup>11</sup>. Outside the binding pocket, DNA intercalating residues stabilize the catalytically active conformation of the DNA by filling the void in the duplex created when the lesion is flipped<sup>13</sup>.

We previously determined that the alkylpurine DNA glycosylase AlkD from *B. cereus*<sup>14</sup> uniquely recognizes damaged DNA without a nucleobase binding pocket or an intercalating residue<sup>15,16</sup>. However, the modified nucleotides in these structures were not in contact with the protein, leaving us to only speculate on the catalytic mechanism. To determine how AlkD excises positively charged substrates without a base binding pocket or an intercalating residue, we determined a new AlkD crystal structure with DNA containing 3-deaza-3-methyladenine (3d3mA), a comparatively stable isostere of 3mA (Fig. 1b and Extended Data Table 1). As in the previous AlkD–DNA complexes, the DNA duplex is bound by the concave surface of the protein. The helix is bent by 30° away from the enzyme while the minor groove surrounding the lesion is widened by 4 Å (Fig. 1c). This distortion induces an equilibrium in which the 3d3mA•T base pair is in either a Watson–Crick or highly sheared conformation. Shearing displaces the 3d3mA nucleobase by 4 Å into the minor groove and towards AlkD, but leaves the base partially stacked in the duplex (Extended Data Fig. 1). In this conformation, the deoxyribose of 3d3mA is in contact with three residues—Trp109, Asp113 and Trp187 (Fig. 1d)—all of which are crucial for lesion excision and are invariant in the AlkD family<sup>15–18</sup>. The carboxylate and two indole side chains cradle the backbone of the lesion and the two flanking nucleotides. Asp113 is in line with the N-glycosidic bond of 3d3mA and thus ideally positioned to stabilize developing positive charge on the deoxyribose as the nucleobase is excised (Fig. 1d). This arrangement also allows Asp113 to position and deprotonate the nucleophilic water for subsequent addition to the oxocarbenium intermediate. Similar interactions between catalytic carboxylate residues and the deoxyribose are achieved by other DNA glycosylases, but only after the lesion has been flipped into the nucleobase binding pocket. By contrast, hypothetical rotation of the 3d3mA nucleotide in the AlkD complex would disrupt these catalytic contacts. The lack of base flipping, however, precludes a general acid from gaining access to protonate the nucleobase substrate, which limits AlkD to excision of cationic lesions.

Unlike 3mA, 3d3mA is uncharged and therefore refractory to depurination at physiological pH<sup>10</sup>. To our surprise, the electron density in the new AlkD–DNA complexes revealed a mixture of intact 3d3mA nucleotide together with AP site and free 3d3mA nucleobase, indicating that cleavage of the N-glycosidic bond had occurred (Fig. 2a and Extended Data Fig. 2). By flash-freezing crystals at various times and determining their structures, we were able to visualize the glycosylase reaction at 1.4–2.0 Å resolution, starting with intact

<sup>1</sup>Department of Biological Sciences and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37232, USA. <sup>2</sup>Department of Chemistry, University of California, Davis, California 95616, USA. <sup>3</sup>Biotechnology Research Center, Toyama Prefectural University, 5180 Kurokawa, Imizu, Toyama 939-0398, Japan.

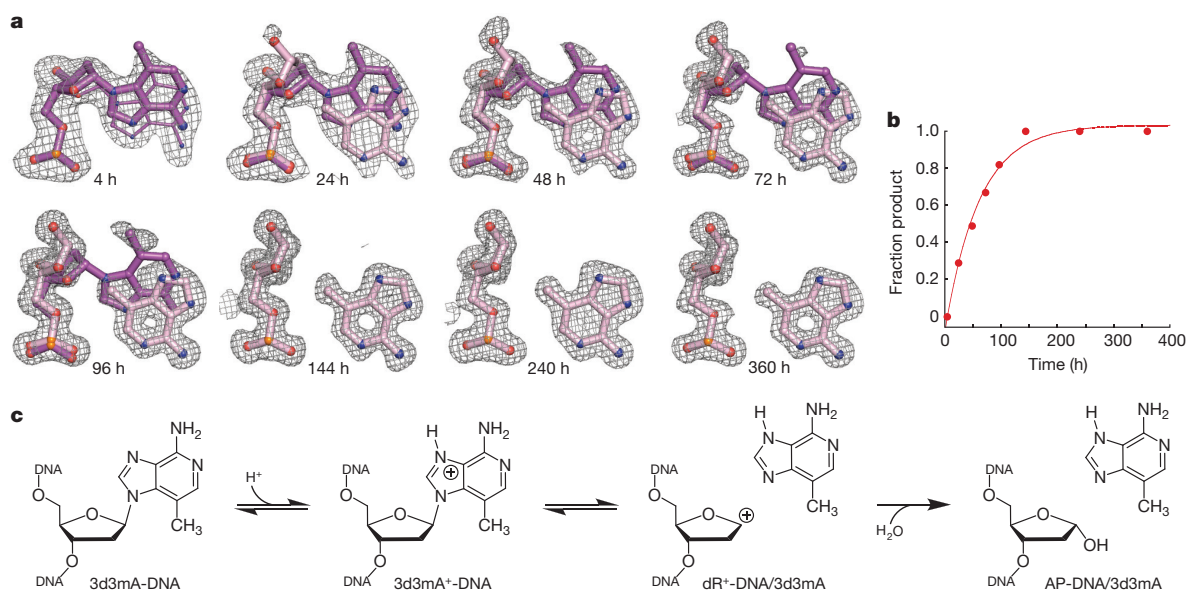


**Figure 1 | Crystallographic reconstruction of the reaction trajectory for 3mA excision.** **a**, Proposed reaction scheme illustrating stepwise depurination of 3mA-DNA. **b**, Corresponding structural analogues used for crystallography. **c**, AlkD–3d3mA-DNA complex. **d**, Substrate-like complex containing 3d3mA-DNA. Watson–Crick (56%) and sheared (44%) conformations of the 3d3mA•T base pair are represented with thin and

thick bonds, respectively. **e**, Intermediate-like complex containing 1aR-DNA and 3mA nucleobase. **f**, Product-like complex containing THF-DNA and 3mA nucleobase. AlkD is coloured blue, the lesion is purple, the opposing thymine is green and the flanking nucleotides are yellow. Charge–dipole and hydrogen-bonding interactions are shown as dashed lines. Annealed omit electron density in **d–f** is contoured to  $2.5\sigma$ .

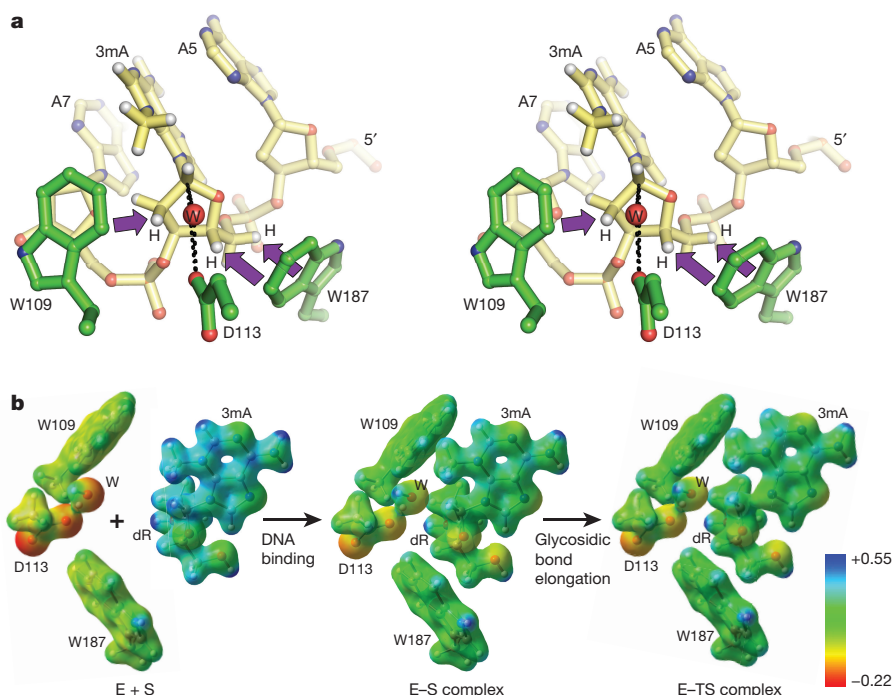
3d3mA-DNA substrate and ending with AP-DNA product (Fig. 2a and Extended Data Tables 1 and 2). Quantifying the fractional occupancies of substrate and product over the course of the reaction gave a rate constant for *in crystallo* base excision of  $4.6 \times 10^{-6} \text{ s}^{-1}$  (Fig. 2b). For comparison, cationic 3mA lesions are excised by AlkD at least 800-fold more rapidly<sup>19</sup>. The unexpected excision of 3d3mA can be

explained by pH-dependent protonation at N7, which would confer positive charge on 3d3mA, activating it for excision by AlkD (Fig. 2c). Given the moderately acidic (pH 5.7) crystallization buffer and the calculated  $\text{pK}_a$  (3.8) of 3d3mA, ~1% of the lesion should be protonated, consistent with the slow rate of *in crystallo* cleavage. By contrast, we did not observe excision of 3d3mA in AlkD–3d3mA-DNA



**Figure 2 | Crystallographic snapshots of 3d3mA excision by AlkD.** **a**, Enzymatic conversion of 3d3mA-DNA substrate (purple) to AP-DNA product and 3d3mA nucleobase (pink). The excised 3d3mA nucleobase is rotated by  $180^\circ$  relative to its position in the non-hydrolysed substrate.

Annealed omit electron density is contoured to  $2.5\sigma$ . **b**, Time course of 3d3mA excision determined from fractional occupancies of substrate and product in the crystal structures. **c**, Proposed reaction scheme showing protonation and excision of 3d3mA.



**Figure 3 | 3mA recognition and excision through charge-dipole and CH- $\pi$  interactions.** **a**, Simulated AlkD-3mA-DNA complex (stereo image). Charge-dipole and hydrogen-bonding interactions are depicted as black dashed lines. CH- $\pi$  interactions are shown as purple arrows. **b**, Electrostatic potential surfaces of catalytic residues and 3mA nucleoside before

protein-DNA binding (E + S), in the enzyme-substrate (E-S) complex and in an approximate enzyme-transition state (E-TS) complex. All structures were determined computationally. Potentials were scaled to  $-0.22$ – $0.55$  atomic units on an isodensity surface of  $0.05$  electrons  $\text{bohr}^{-3}$ .

complexes crystallized at pH 7.0 (Extended Data Fig. 3 and Extended Data Table 3). We also did not observe cleavage in crystals grown at pH 5.7 in which AlkD bound 3d3mA-DNA in a non-catalytic orientation that placed the lesion on the opposite face of the duplex and away from the protein<sup>16</sup> (Extended Data Fig. 4). Thus, the AlkD-3d3mA-DNA structure presented here represents a bona fide enzyme-substrate complex that enabled visualization of the endpoints of the glycosylase reaction.

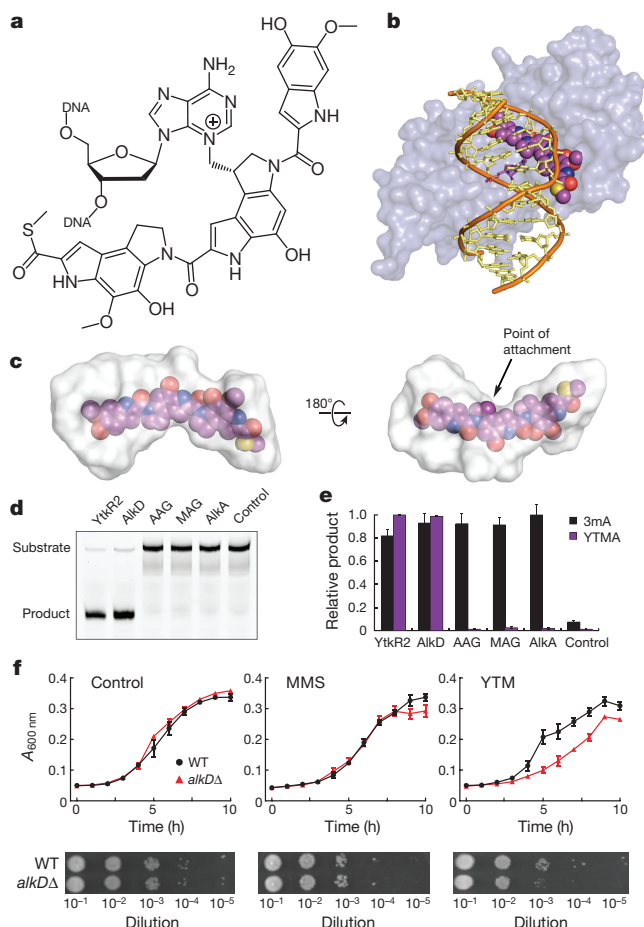
We probed the intervening step of the reaction trajectory by determining a structure representing the oxocarbenium intermediate, using DNA containing 1'-aza-2',4'-dideoxyribose (1aR) and 3mA nucleobase (Fig. 1b and Extended Data Table 3). Relative to the position of the 3d3mA nucleotide, the cationic 1aR moiety is shifted slightly towards the surface of AlkD, which enhances electrostatic interactions with Asp113 and the nucleophilic water (Fig. 1e). These same interactions would stabilize the high-energy oxocarbenium intermediate formed upon cleavage of the glycosidic bond. A nearly identical arrangement is present in the product-like complex containing tetrahydrofuran (THF)-DNA and 3mA nucleobase (Fig. 1b, f and Extended Data Table 3). The only notable exception is a small rotation of the neutral THF ring away from Asp113. In both ternary complexes, the 3mA nucleobase is retained in the DNA duplex and paired with the complementary thymine, maintaining stacking interactions with the flanking bases (Fig. 1e, f). While base stacking is altered upon shearing of the 3d3mA•T base pair, it is never fully disrupted and is completely restored following cleavage of the *N*-glycosidic bond. In stark contrast to the traditional base-flipping mechanism, there is no evidence from these structures that a void in the duplex is created at any point along the reaction trajectory that would require stabilization by a DNA intercalating residue. Furthermore, these structures indicate that after shearing of the 3d3mA•T base pair, minimal movement of the protein and the DNA is necessary for either bond cleavage or nucleophile addition to occur.

The manner in which Trp109 and Trp187 contribute to binding and catalysis<sup>17,18</sup> has so far been unknown. The new AlkD-DNA

structures reveal that both residues form CH- $\pi$  interactions<sup>20</sup> with C2', C4' and C5' of the 3d3mA nucleotide (Fig. 3a). To gain insight into how these weakly polar contacts contribute to recognition and excision of 3mA, we used a reductive computational approach with only the three catalytic residues, the nucleophilic water and the lesion. Electrostatic potential calculations showed that substantial cationic character is present on both the 3mA nucleobase and the deoxyribose (Fig. 3b and Extended Data Fig. 5). Importantly, increased positive charge on the deoxyribose correlated with stronger calculated binding energies, suggesting a means by which AlkD might recognize cationic alkylpurine lesions through backbone contacts (Extended Data Fig. 5). This would enhance detection of altered base pairing or base stacking<sup>21</sup> without the need to interact with the modified nucleobase directly. The electrostatic potential calculations also revealed that additional positive charge is transferred to the deoxyribose upon elongation of the *N*-glycosidic bond (Extended Data Fig. 5). Correspondingly, preferential binding and stabilization of this transition-state-like structure resulted in a theoretical  $10^3$ – $10^4$ -fold rate enhancement of glycosidic bond cleavage. Roughly half of this enhancement was attributed to the increasingly polar CH- $\pi$  interactions provided by Trp109 and Trp187 (Extended Data Fig. 5). While interactions of this type are widespread among proteins and are prevalent in protein-DNA complexes<sup>22,23</sup>, to our knowledge, this is the first indication that CH- $\pi$  interactions might function in a catalytic capacity in DNA repair. These interactions are reminiscent of the  $\pi$ - $\pi$  and cation- $\pi$  interactions used by base-flipping enzymes to recognize extrahelical nucleobases<sup>24,25</sup>, but are distinctly different in their involvement of the deoxyribose, and may be fundamental to lesion excision in the absence of base flipping.

The unique mechanism reported here explains our previous finding that AlkD, but not human AAG, removes bulky, cationic pyridyl-oxobutyl adducts<sup>16</sup>. While extrahelical pyridyl-oxobutyl adducts are accommodated by the large nucleobase binding pockets of alkyltransferase-like proteins<sup>26</sup>, the relatively small active site pockets of AAG and other base-flipping DNA glycosylases impose tighter steric limitations. AlkD, however, does not rely on a base-flipping mechanism





**Figure 4 | Excision of N3-yatakemycyladenine by AlkD.** **a**, YTMA. **b**, Modelled AlkD-YTMA-DNA complex. **c**, Solvent-filled cavity between AlkD and 3d3mA-DNA into which YTM was modelled. **d**, In vitro YTMA excision monitored by separation of full-length YTMA-DNA substrate from cleaved AP-DNA product by denaturing gel electrophoresis. **e**, Quantification of in vitro excision of 3mA (black) and YTMA (purple) by diverse alkylpurine DNA glycosylases. Error bars denote s.d. from three replicate experiments. **f**, Growth of wild-type (black) and *alkD*-knockout (red) strains of *B. anthracis* in the presence of no drug (control), 2 mM MMS and 10 nM YTM. Error bars denote s.e.m. from three replicate experiments.

and therefore is not restricted in this fashion, suggesting it may have a cellular role in the repair of bulky lesions. Specificity for bulky alkyl adducts would explain the coexistence of several alkylpurine glycosylases in numerous bacterial species<sup>14</sup>. Consistent with this notion, an AlkD orthologue from *Streptomyces* sp. TP-A0356 has been found to excise 815-dalton (Da) N3-yatakemycyladenine (YTMA) adducts from DNA<sup>27</sup> (Fig. 4a). Yatakemycin (YTM) is a minor-groove DNA-alkylating agent belonging to the duocarmycin family of antibiotic and antitumour drugs<sup>6</sup>. The gene cluster responsible for biosynthesis of YTM encodes the DNA glycosylase YtkR2, which confers resistance to YTM by excising YTMA lesions<sup>27</sup>.

On the basis of the sequence similarity between YtkR2 and AlkD, we proposed that AlkD would also have activity for YTMA. There is an extended solvent-filled volume between the protein and the minor groove in the AlkD-3d3mA-DNA complex that can accommodate YTM modelled onto the C3 position of 3d3mA without introducing steric clashes or disrupting catalytic interactions (Fig. 4b, c). Using a standard in vitro DNA glycosylase assay, we found that AlkD excised YTMA from DNA with the same efficiency as YtkR2 (Fig. 4d, e). By contrast, the alkylpurine DNA glycosylases AAG, MAG and AlkA failed to excise YTMA but readily removed 3mA (Fig. 4d, e). To determine the specificity of AlkD for YTMA in cells, we constructed a *Bacillus anthracis* strain lacking *alkD* and tested its sensitivity against YTM and

methyl methanesulfonate (MMS). MMS primarily produces 3mA and N7-methylguanine adducts that are removed by canonical alkylpurine DNA glycosylases<sup>5</sup>. In the absence of alkylating agent, deletion of *alkD* had no effect on the growth of *B. anthracis* (Fig. 4f and Extended Data Fig. 6). Similarly, *alkD* cells were no more sensitive to MMS than wild-type cells (Fig. 4f and Extended Data Fig. 6), most likely as a result of activity from the other alkylpurine glycosylases (AAG, AlkA and AlkC) still present in the deletion strain. Conversely, deletion of *alkD* caused a marked increase in sensitivity to YTM, consistent with AlkD-catalysed excision of YTMA in vivo (Fig. 4f and Extended Data Fig. 6). This suggests that the primary role of AlkD in the cell is to repair larger alkyl adducts such as YTMA that are not normally corrected by the base excision repair pathway.

This work establishes that, contrary to dogma, substrate recognition and catalysis by DNA glycosylases can occur in the absence of base flipping. There is evidence that other enzymes are able to detect DNA damage and even discriminate between different chemical modifications before flipping<sup>28–30</sup>. That AlkD is limited to removing inherently labile cationic lesions strongly suggests that the main function of the nucleobase binding pocket in canonical DNA glycosylases is to increase the leaving-group potential of the nucleobase substrate, as opposed to merely discriminating against non-substrate nucleobases. With this capability, however, comes a limit on the size of the adduct that can be excised. Ultimately, the greatest benefit of a non-base-flipping mechanism may be the ability to repair bulky lesions.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 July; accepted 18 September 2015.

Published online 28 October 2015.

- Fromme, J. C. & Verdine, G. L. Base excision repair. *Adv. Protein Chem.* **69**, 1–41 (2004).
- Hitomi, K., Iwai, S. & Tainer, J. A. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst.)* **6**, 410–428 (2007).
- Slupphaug, G. et al. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87–92 (1996).
- Stivers, J. T. Site-specific DNA damage recognition by enzyme-induced base flipping. *Prog. Nucleic Acid Res. Mol. Biol.* **77**, 37–65 (2004).
- Brooks, S. C., Adhikary, S., Robinson, E. H. & Eichman, B. F. Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta* **1834**, 247–271 (2013).
- Igarashi, Y. et al. Yatakemycin, a novel antifungal antibiotic produced by *Streptomyces* sp. TP-A0356. *J. Antibiot. (Tokyo)* **56**, 107–113 (2003).
- Alberts, S. R., Suman, V. J., Pitot, H. C., Camoriano, J. K. & Rubin, J. Use of KW-2189, a DNA minor groove-binding agent, in patients with hepatocellular carcinoma: a north central cancer treatment group (NCCTG) phase II clinical trial. *J. Gastrointest. Cancer* **38**, 10–14 (2007).
- Friedberg, E. C. et al. *DNA Repair and Mutagenesis* 2nd edn (ASM Press, 2006).
- Larson, K., Sahm, J., Shenkar, R. & Strauss, B. Methylation-induced blocks to in vitro DNA replication. *Mutat. Res.* **150**, 77–84 (1985).
- Plosky, B. S. et al. Eukaryotic Y-family polymerases bypass a 3-methyl-2'-deoxyadenosine analog in vitro and methyl methanesulfonate-induced DNA damage in vivo. *Nucleic Acids Res.* **36**, 2152–2162 (2008).
- Drohat, A. C. & Maiti, A. Mechanisms for enzymatic cleavage of the N-glycosidic bond in DNA. *Org. Biomol. Chem.* **12**, 8367–8378 (2014).
- Stivers, J. T. & Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **103**, 2729–2760 (2003).
- Hendershot, J. M. & O'Brien, P. J. Critical role of DNA intercalation in enzyme-catalyzed nucleotide flipping. *Nucleic Acids Res.* **42**, 12681–12690 (2014).
- Alseth, I. et al. A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol. Microbiol.* **59**, 1602–1609 (2006).
- Rubinson, E. H., Metz, A. H., O'Quin, J. & Eichman, B. F. A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD. *J. Mol. Biol.* **381**, 13–23 (2008).
- Rubinson, E. H., Gowda, A. S., Spratt, T. E., Gold, B. & Eichman, B. F. An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* **468**, 406–411 (2010).
- Dalhous, B. et al. Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats. *Nucleic Acids Res.* **35**, 2451–2459 (2007).
- Mullins, E. A., Robinson, E. H. & Eichman, B. F. The substrate binding interface of alkylpurine DNA glycosylase AlkD. *DNA Repair (Amst.)* **13**, 50–54 (2014).



19. Mullins, E. A. *et al.* An HPLC-tandem mass spectrometry method for simultaneous detection of alkylated base excision repair products. *Methods* **64**, 59–66 (2013).
20. Plevin, M. J., Bryce, D. L. & Boisbouvier, J. Direct detection of CH/ $\pi$  interactions in proteins. *Nature Chem.* **2**, 466–471 (2010).
21. Yang, W. Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair (Amst.)* **5**, 654–666 (2006).
22. Brandl, M., Weiss, M. S., Jabs, A., Suhnel, J. & Hilgenfeld, R. C-H $\cdots\pi$ -interactions in proteins. *J. Mol. Biol.* **307**, 357–377 (2001).
23. Wilson, K. A., Kellie, J. L. & Wetmore, S. D. DNA-protein  $\pi$ -interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. *Nucleic Acids Res.* **42**, 6726–6741 (2014).
24. Metz, A. H., Hollis, T. & Eichman, B. F. DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG). *EMBO J.* **26**, 2411–2420 (2007).
25. Wilkinson, O. J. *et al.* Alkyltransferase-like protein (At11) distinguishes alkylated guanines for DNA repair using cation- $\pi$  interactions. *Proc. Natl Acad. Sci. USA* **109**, 18755–18760 (2012).
26. Tubbs, J. L. *et al.* Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* **459**, 808–813 (2009).
27. Xu, H. *et al.* Self-resistance to an antitumor antibiotic: a DNA glycosylase triggers the base-excision repair system in yatakemycin biosynthesis. *Angew. Chem. Int. Ed. Engl.* **51**, 10532–10536 (2012).
28. Qi, Y. *et al.* Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* **462**, 762–766 (2009).
29. Imamura, K., Averill, A., Wallace, S. S. & Doublet, S. Structural characterization of viral ortholog of human DNA glycosylase NEIL1 bound to thymine glycol or 5-hydroxyuracil-containing DNA. *J. Biol. Chem.* **287**, 4288–4298 (2012).
30. Adhikary, S. & Eichman, B. F. Analysis of substrate specificity of *Schizosaccharomyces pombe* Mag1 alkylpurine DNA glycosylase. *EMBO Rep.* **12**, 1286–1292 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank C. Rizzo and T. Johnson Salyard for assistance with oligodeoxynucleotide synthesis and E. Skaar and L. Mike for assistance with genetic manipulation of *B. anthracis*. This work was funded by the National Science Foundation (MCB-1122098 and MCB-1517695 to B.F.E.) and the National Institutes of Health (R01ES019625 to B.F.E. and R01CA067985 to S.S.D.). Support for the Vanderbilt Robotic Crystallization Facility was also provided by the National Institutes of Health (S10RR026915). Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy Office of Science by Argonne National Laboratory, was supported by the US Department of Energy (DE-AC02-06CH11357). Use of LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (085P1000817). E.A.M. and Z.D.P. were partially supported by the Vanderbilt Training Program in Environmental Toxicology (T32ES07028).

**Author Contributions** E.A.M. and B.F.E. conceived the project; E.A.M., R.S., Z.D.P. and B.F.E. designed experiments; E.A.M. performed biochemical and structural experiments; R.S. performed cellular experiments; Z.D.P. performed computational experiments; P.K.Y., S.S.D. and Y.I. supplied reagents; E.A.M., R.S., Z.D.P. and B.F.E. analysed data and wrote the paper; all authors commented on the manuscript.

**Author Information** Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 5CL3, 5CL4, 5CL5, 5CL6, 5CL7, 5CL8, 5CL9, 5CLA, 5CLB, 5CLC, 5CLD and 5CLE. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.F.E. ([brandteichman@vanderbilt.edu](mailto:brandteichman@vanderbilt.edu)).

## METHODS

**Oligodeoxynucleotide synthesis.** 3d3mA phosphoramidite was purchased from Berry and Associates. 1aR phosphoramidite was synthesized with minor modification of a previously described method<sup>31</sup>. Both were incorporated into oligodeoxynucleotides using standard solid-phase techniques. The resulting products were purified by reverse-phase HPLC and verified by mass spectrometry. All other oligodeoxynucleotides were purchased from Integrated DNA Technologies and used without further purification.

**Protein purification.** Human ( $\Delta 79$ )AAG<sup>32</sup>, *Saccharomyces cerevisiae* MAG<sup>30</sup>, *Escherichia coli* AlkA<sup>33</sup> and *B. cereus* AlkD<sup>15</sup> were purified as previously described. The gene encoding YtkR2 (GenBank accession IADZ13541) was synthesized by DNA2.0 and ligated into a modified pET27 expression vector encoding a Rhinovirus 3C cleavable hexahistidine tag. YtkR2 overproduction in *E. coli* HMS174(DE3) cells was induced at 16 °C after addition of 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). Cells were collected from LB medium by centrifugation, resuspended in buffer L (50 mM Tris-HCl, pH 7.5, 500 mM NaCl and 20% (v/v) glycerol) and lysed on ice by gentle sonication. Cleared lysate was applied to a Ni-NTA column equilibrated in buffer L. The column was then washed with buffer L containing 2 mM histidine and eluted with buffer L containing 100 mM EDTA. Pooled fractions were supplemented with 2 mM dithiothreitol (DTT) before overnight cleavage of the affinity tag. Cleaved YtkR2 was diluted tenfold in buffer H (50 mM Tris-HCl, pH 7.5, 20% (v/v) glycerol, 2 mM DTT and 0.1 mM EDTA) before being applied to a heparin Sepharose column equilibrated in buffer H. The column was then washed with buffer H containing 50 mM NaCl and eluted by linearly increasing to buffer H containing 1 M NaCl. Pooled fractions were passed through a Ni-NTA column equilibrated in buffer L to remove trace protein contaminants. The column was subsequently rinsed with buffer L containing 2 mM histidine to elute weakly bound YtkR2. The flow-through and the rinse were then combined, concentrated by ultracentrifugation and applied to a Superdex 200 column equilibrated in buffer S (20 mM Bis-tris propane, pH 6.5, 400 mM NaCl, 20% (v/v) glycerol, 2 mM DTT and 0.1 mM EDTA). YtkR2 was eluted with additional buffer S, concentrated to 3 mg ml<sup>-1</sup> by ultracentrifugation, flash-frozen in liquid nitrogen and stored at -80 °C.

**Crystallization of AlkD-DNA.** DNA annealing reactions contained 0.54 mM 9-mer A [d(AAGCAXACC)/d(TGGTTTGCT)], 9-mer B [d(AAGCCXCCC)/d(TGGGTGGCT)], or 12-mer [d(CCCGAXAGTCCG)/d(CGGACTTTCGGG)] oligodeoxynucleotides, 10 mM MES, pH 6.5, and 40 mM NaCl. Reactions containing 1aR•T-DNA or THF•T-DNA were supplemented with 27 mM 3mA nucleobase. Strands were annealed by heating to 85 °C and slowly cooling to 20 °C over several hours. AlkD-DNA complexes were formed by mixing 0.45 mM protein and 0.54 mM DNA solutions in equal volumes and incubating at 4 °C for 30 min. Complexes containing 9-mer DNA were crystallized using the sitting drop vapour diffusion method. Drops were prepared from 2  $\mu$ l of protein-DNA solution (0.22 mM AlkD and 0.27 mM DNA), 2  $\mu$ l of reservoir solution (22–25% (w/v) PEG 8,000, 50 mM HEPES, pH 7.0, and 50 mM CaCl<sub>2</sub>), and 1  $\mu$ l of additive solution (5% (w/v) benzamidine hydrochloride) and equilibrated against an additional 500  $\mu$ l of reservoir solution at 21 °C. Crystals were collected after 24 h, briefly soaked in reservoir solution supplemented with 15% (v/v) glycerol and flash-cooled in liquid nitrogen. Complexes containing 12-mer DNA were crystallized using the hanging drop vapour diffusion technique. Drops were assembled from 2  $\mu$ l of protein-DNA solution (0.22 mM AlkD and 0.27 mM DNA), 2  $\mu$ l of reservoir solution (15–19% (w/v) PEG 4,000, 42 mM sodium acetate, pH 4.6, 85 mM ammonium acetate and 5% (v/v) glycerol) and 1  $\mu$ l of seed solution (submicroscopic crystals of AlkD-1mA•T-DNA) and equilibrated against an additional 500  $\mu$ l of reservoir solution at 21 °C. Crystals were collected after 4–360 h, briefly soaked in reservoir solution supplemented with 15% (v/v) glycerol and flash-cooled in liquid nitrogen.

**X-ray data collection and structure refinement.** X-ray diffraction data were collected at beamlines 21-ID-F and 21-ID-G at the Advanced Photon Source and processed using HKL2000 (ref. 34). Data collection statistics are provided in Extended Data Tables 1–3. A previously determined model of AlkD (PDB accession 3BVS) was positioned with Phaser<sup>35</sup>, while DNA was manually built in Coot<sup>36</sup>, guided by inspection of 2mF<sub>o</sub> - DF<sub>c</sub> and mF<sub>o</sub> - DF<sub>c</sub> electron density maps. The entirety of the 12-mer oligodeoxynucleotide duplex was readily apparent, as were AlkD residues 1–229 and two non-native residues (-1–0) from the cleaved N-terminal affinity tag, but not the last eight residues (230–237) at the C terminus. As with the 12-mer, the complete 9-mer duplexes were visible in the density maps. However, three residues (52–54) between helix C and helix D and twelve residues (226–237) at the C terminus could not be reliably modelled. Atomic coordinates, temperature factors and fractional occupancies were refined in PHENIX<sup>37</sup>. The final AlkD-DNA models were validated using MolProbity<sup>38</sup> and contained no residues in the disallowed regions of the Ramachandran plot. Refinement and validation statistics are given in Extended Data Tables 1–3.

All structure images were created in PyMOL (<https://www.pymol.org>). mF<sub>o</sub> - DF<sub>c</sub> omit electron density maps were calculated using PHENIX by removing the lesion and the opposing thymine and then performing simulated annealing on the remaining AlkD-DNA complex to minimize model bias. Maps were carved around the omitted atoms with a 1.5–2.0 Å radius and contoured to 2.5 $\sigma$ . YTM was manually docked in the AlkD-3d3mA-DNA complex after defining the cavity between the protein and the DNA with Hollow<sup>39</sup>. Simulated structures shown in Fig. 3a and Extended Data Fig. 1c were generated from AlkD-3d3mA-DNA crystal structures by manually removing or transmuting atoms and without performing subsequent computational optimization.

**Preparation of YTM-DNA.** YTM was purified from *Streptomyces* sp. TP-A0356 culture as previously described<sup>6</sup>. Addition reactions containing 150  $\mu$ M YTM, 10  $\mu$ M fluorescein (FAM)-labelled DNA (FAM-d(CGGGCGGCGGCAAGGGCGCGGGCC)/d(GGCCCCGCGCCCTTTGCCGCGCCCCG); underline denotes nucleotide modified by YTM), 10 mM MES, pH 6.5, 40 mM NaCl and 10% (v/v) DMSO were incubated at 25 °C for 24 h and periodically mixed by inversion. YTM-modified DNA was then separated from free YTM by passage through a G25 size exclusion column equilibrated in 10 mM MES, pH 6.5, and 40 mM NaCl and stored at -80 °C.

**Quantification of base excision.** Reaction mixtures containing 5  $\mu$ M enzyme, 10  $\mu$ g of methylated calf thymus DNA or 100 nM YTM-modified oligodeoxynucleotide duplex, 50 mM Bis-tris propane, pH 6.5, 100 mM NaCl, 5% (v/v) glycerol, 2 mM DTT, 0.1 mM EDTA and 0.1 mg ml<sup>-1</sup> BSA were incubated at 25 °C for 6 h. Excision of 3mA was quantified by measuring cleaved nucleobase using a previously described HPLC-MS/MS method<sup>19</sup>. Excision of YTM was quantified by separating the 25-mer substrate from the 12-mer product using standard electrophoretic techniques<sup>19</sup>. Strand breakage of 25-mer abasic oligodeoxynucleotide was induced by heating the reaction mixture at 70 °C for 30 min in the presence of 0.2 M NaOH.

**Preparation of alkD cells.** The 1-kilobase flanking regions surrounding *alkD* were inserted into the knockout-plasmid pLM4 using standard molecular biology techniques<sup>40</sup>. The modified plasmid was propagated in *E. coli* K1077 before introduction by electroporation into *B. anthracis* Sterne cells<sup>41,42</sup>. *B. anthracis* colonies containing the plasmid were grown on LB plates supplemented with 20  $\mu$ g ml<sup>-1</sup> kanamycin at 42.5 °C for 1–2 days to generate merodiploids with plasmid DNA integrated into their chromosomal DNA. Merodiploids were then grown in LB medium for 1 day at 30 °C to facilitate elimination of redundant DNA from their genomes. Cultures were serially diluted and grown on LB plates without kanamycin for 1 day at 30 °C so that colonies lacking *alkD* could be identified by PCR screening.

**Determination of YTM and MMS resistance.** Overnight cultures of *B. anthracis* Sterne (wild-type and *alkD* $\Delta$ ) grown at 30 °C were diluted 1:100 in 100  $\mu$ l of LB medium in the presence or absence of 2–40 nM YTM or 1–10 mM MMS in a 96-well flat-bottom plate. The plate was incubated at 30 °C with shaking for 20 h, and cell density was measured at 600 nm every hour using a Synergy 2 multi-detector microplate reader. Before each measurement, the plate was gently vortexed to ensure full resuspension of sedimented cells. Experiments were performed in triplicate.

Overnight cultures of *B. anthracis* Sterne (wild-type and *alkD* $\Delta$ ) grown at 30 °C were diluted 1:100 in 5 ml of LB medium and incubated at 30 °C with shaking until early logarithmic phase. Culture aliquots (5  $\mu$ l) were then tenfold serially diluted (1:10<sup>-1</sup>–1:10<sup>-5</sup>) and spotted on LB plates prepared with or without 2–40 nM YTM or 1–10 mM MMS. Plates were incubated at 37 °C and imaged after 2 days. Experiments were performed in duplicate.

**Calculation overview.** All calculations were performed using Gaussian 09 ([http://www.gaussian.com/g\\_prod/g09.htm](http://www.gaussian.com/g_prod/g09.htm)). Atomic coordinates for computationally optimized structures are provided in Supplementary Data 1.

**pK<sub>a</sub> calculation.** The aqueous pK<sub>a</sub> microacidity constant of N7-protonated 3-deaza-3,9-dimethyladenine (3d3m9mA) was calculated using the isodesmic reaction method<sup>43</sup> with N7-protonated 9-methyladenine (9mA) used as the reference acid (aqueous pK<sub>a</sub> = 2.96)<sup>44</sup>. Neutral and N7-protonated 9mA and 3d3m9mA were optimized to minima using Truhlar's M06-2X density functional<sup>45</sup> with the split-valence 6-311+G(2d,p) basis set and the SMD solvation model<sup>46</sup>. Optimized structures were confirmed minima by vibrational frequency analyses. Single point evaluations were carried out on the optimized structures in the presence of implicit solvent (water) at the M06-2X/aug-cc-pVTZ level. Thermal corrections to Gibbs free energies were taken from the frequency calculations and applied to electronic energies determined at the M06-2X/aug-cc-pVTZ level. Proton exchange energetics ( $\Delta G_{\text{exch}}^{\text{chrg}}$ ) between 9mA and 3d3m9mA were then determined using the corrected electronic energies.

**Constrained optimization of AlkD-lesion structures.** The crystal structure of AlkD bound to 3d3mA-DNA served as the starting point for all computational structural optimizations. Atomic coordinates were extracted for AlkD residues

Trp109, Asp113 and Trp187 as well as HOH308 and 3d3mA. Hydrogen atoms were added manually to satisfy all valences, and Asp113 was assumed to be in carboxylate form. To reduce computational cost during structural optimizations and subsequent calculations, protein residues were truncated to their respective side chains ( $\beta$ -carbons as methyl groups), and 3d3mA was truncated to the corresponding 2'-deoxynucleoside. The starting point for the structure of 3mA was generated by replacing carbon with nitrogen at the 3-position of the purine ring in 3d3mA. The starting point for the transition state approximation (TSA) was generated by elongating the glycosidic bond of optimized 3mA by 0.5 Å. This distance was chosen based on the computationally determined transition state for excision of uracil by UDG<sup>47</sup>. Finally, the starting point for dR<sup>+</sup> was generated by deleting the 3d3mA nucleobase from the corresponding 2'-deoxynucleoside, leaving an open valence on C1'.

Structural optimizations were performed in the gas phase using the M06-2X functional in conjunction with the 6-31+G(d) basis set. In all cases, the Cartesian coordinates for all non-hydrogen atoms in AlkD residues Trp109, Asp113 and Trp187 as well as HOH308 were frozen using the freeze code -1 in the molecular coordinates specification. Additionally, the 3'- and 5'-oxygen atoms of all 2'-deoxynucleosides were frozen to simulate the experimentally observed lack of motion in the phosphodiester backbone in the full AlkD-DNA complexes. All other atoms, with the exception of C1' and N9 in the TSA, were allowed to move during optimizations. Stationary points derived from geometry optimizations on these reduced-dimension potential energy surfaces were verified as minima by vibrational frequency analyses. Frequency analyses were performed at the same level of theory and on the same reduced-dimension surfaces as the geometry optimizations. Using this approach, freezing Cartesian coordinates results in zeroing of elements in the Hessian associated with the frozen atoms. As such, the expected imaginary frequency resulting from elongation of the glycosidic bond in the TSA was not observed. All structures, including the TSA, afforded zero imaginary frequencies.

**Calculation of binding energy and catalytic rate enhancement.** Binding energies of the truncated AlkD residues and the catalytic water to 3d3mA, 3mA, TSA and dR<sup>+</sup> were determined using the M06-2X functional. Structural optimizations were carried out as described above using the 6-31+G(d) basis set, and single point evaluations were carried out using the 6-311++G(3df,2p) basis set. The basis set superposition error was accounted for using the counterpoise method<sup>48</sup>, in which AlkD residues and the catalytic water were defined as the first fragment and the 2'-deoxynucleosides were defined as the second fragment. Because heavy atoms in AlkD residues were frozen at their experimentally determined atomic positions, binding energies were evaluated as differences in zero point-uncorrected electronic energies. This approach reproduced experimental gas-phase binding energies to within 1 kcal mol<sup>-1</sup> for two prototypical CH- $\pi$  systems, including neutral CH<sub>4</sub>/benzene (experimental 1.03–1.13 kcal mol<sup>-1</sup> versus calculated 1.6 kcal mol<sup>-1</sup>)<sup>49</sup> and cationic (CH<sub>3</sub>)<sub>4</sub>N<sup>+</sup>/benzene (experimental 9.4 kcal mol<sup>-1</sup> versus calculated 10.3 kcal mol<sup>-1</sup>)<sup>50</sup>. To determine the individual contributions of Trp109, Asp113, Trp187 and HOH308 to ensemble binding energies, the counterpoise method was again employed as described above, but with the atoms in Trp109, Asp113, Trp187 or HOH308 replaced with the corresponding Gaussian ghost atoms. The contribution of each residue to the total binding energy (defined as BE<sub>individual(x)</sub> = BE<sub>ensemble</sub> - BE<sub>ensemble,ghost(x)</sub>) was computed as the difference between the ensemble binding energy and that where Trp109, Asp113, Trp187 or HOH308 had been replaced with ghost atoms. Complexes containing ghost atoms were not reoptimized, which may affect individual binding energies. However, using this method, between 92% and 99% of the ensemble binding energy was accounted for in each complex.

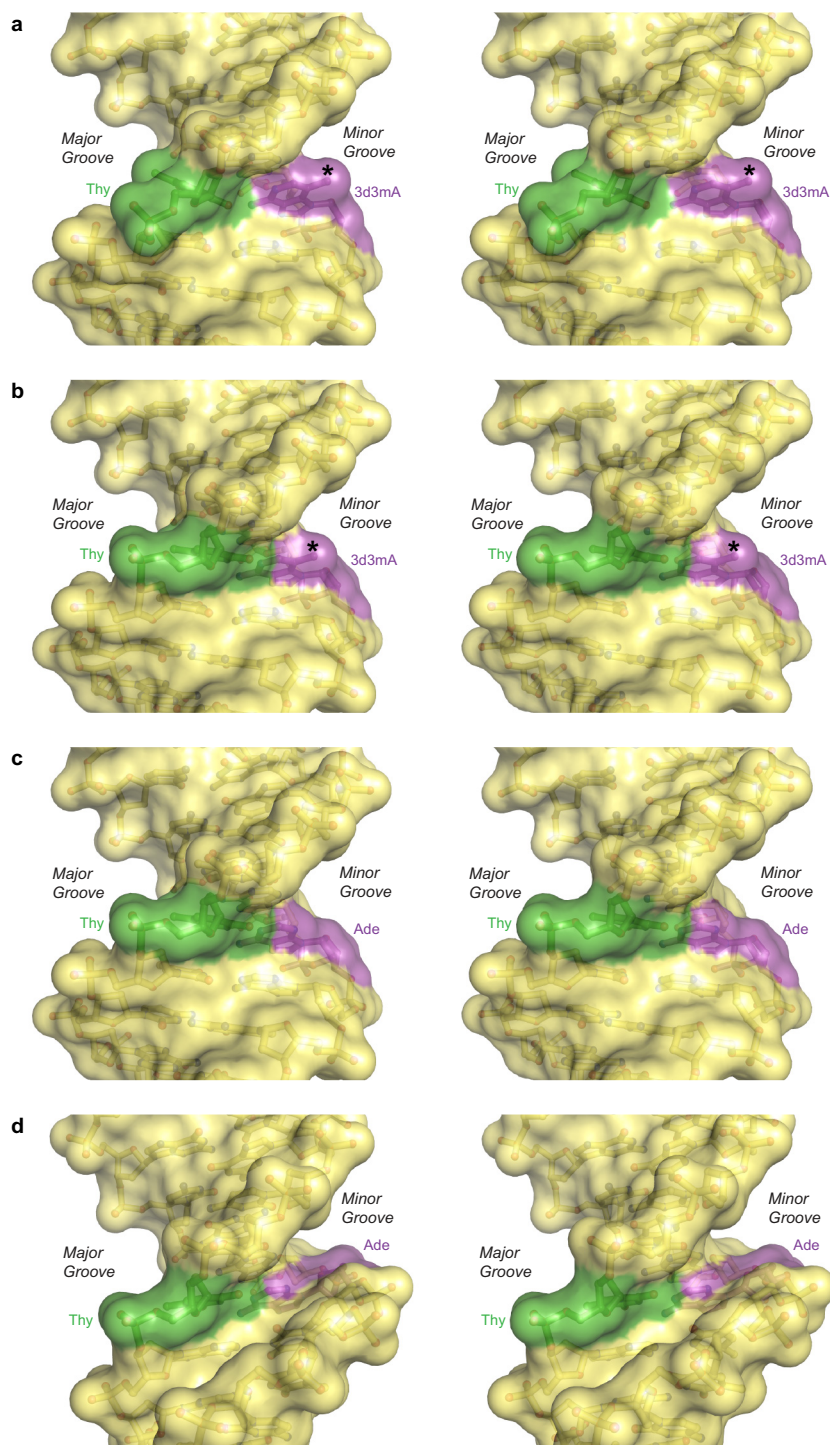
The barrier to bond elongation (defined as  $\Delta E = E_{\text{TSA}} - E_{3\text{mA}}$ ) was calculated in the presence ( $\Delta E = 18.1$  kcal mol<sup>-1</sup>) and absence ( $\Delta E = 23.0$  kcal mol<sup>-1</sup>) of Trp109, Asp113, Trp187 and HOH308. The difference in these values ( $\Delta\Delta E = 4.9$  kcal mol<sup>-1</sup>) was taken as the catalytic contribution of AlkD to 3mA excision. This process was repeated for complexes in which atoms in each residue or the catalytic water had been replaced with ghost atoms. For each series,  $\Delta\Delta E$  values were computed as outlined above, and the difference between each of these  $\Delta\Delta E$  values and that of the full ensemble was taken as the individual contribution of the omitted residue to catalysis. Estimated rate enhancements were calculated from these values by substituting  $\Delta\Delta E$  for  $\Delta\Delta G^\ddagger$  in the Eyring equation at 25 °C. The barrier to bond elongation ( $\Delta E = 23.0$  kcal mol<sup>-1</sup>) calculated with this approach is consistent with the activation enthalpy for depurination of 3mA ( $\Delta H^\ddagger = 23.5$  kcal mol<sup>-1</sup>) extrapolated from empirically determined half-lives<sup>51</sup>.

**Quantitation of charge transfer and generation of electrostatic potential maps.** Charge transfer from AlkD residues Trp109, Asp113 and Trp187 and HOH308 to the lesions was quantitated by performing Merz-Singh-Kollman (MK) population

analyses<sup>52,53</sup> on each complex and on each individual component (either AlkD and the catalytic water or the lesion). To correct for potential basis set superposition error in population analyses, the appropriate Gaussian ghost atoms were used to maintain consistent basis between fragments and complexes. Partial atomic charges derived from MK population analyses were summed for all atoms in AlkD and for the lesion, affording the corresponding group charges. The arithmetic difference between group charges in the AlkD-lesion complex and AlkD or the lesion alone were taken as the amount of charge transferred upon complex formation and expressed as percentages. Electrostatic potential maps were generated from densities computed at the M06-2X/6-311++G(3df,2p) level with ghost atoms omitted for visual clarity. Potentials were scaled to -0.22–0.55 atomic units on an isodensity surface of 0.05 electrons bohr<sup>-3</sup>.

- Chu, A. M., Fetting, J. C. & David, S. S. Profiling base excision repair glycosylases with synthesized transition state analogs. *Bioorg. Med. Chem. Lett.* **21**, 4969–4972 (2011).
- O'Brien, P. J. & Ellenberger, T. Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines. *Biochemistry* **42**, 12418–12429 (2003).
- Bjelland, S., Birkeland, N. K., Benneche, T., Volden, G. & Seeberg, E. DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the AlkA enzyme in *Escherichia coli*. *J. Biol. Chem.* **269**, 30489–30495 (1994).
- Otwowski, Z. & Minor, W. Processing of X-ray diffraction data. *Methods Enzymol.* **276**, 307–326 (1997).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
- Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).
- Mike, L. A. et al. Two-component system cross-regulation integrates *Bacillus anthracis* response to heme and cell envelope stress. *PLoS Pathog.* **10**, e1004044 (2014).
- Sterne, M. Avirulent anthrax vaccine. *Onderstepoort J. Vet. Sci. Anim. Ind.* **21**, 41–43 (1946).
- Stauff, D. L. & Skaar, E. P. *Bacillus anthracis* HssRS signalling to HrtAB regulates haem resistance during infection. *Mol. Microbiol.* **72**, 763–778 (2009).
- Ho, J. & Coote, M. L. A universal approach for continuum solvent pK<sub>a</sub> calculations: are we there yet? *Theor. Chem. Acc.* **125**, 3–21 (2010).
- Kapinos, L. E., Opershall, B. P., Larsen, E. & Sigel, H. Understanding the acid-base properties of adenosine: the intrinsic basicities of N1, N3 and N7. *Chemistry* **17**, 8156–8164 (2011).
- Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241 (2008).
- Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).
- Dinner, A. R., Blackburn, G. M. & Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **413**, 752–755 (2001).
- Boys, S. F. & Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **19**, 553–566 (1970).
- Shibasaki, K., Fujii, A., Mikami, N. & Tsuzuki, S. Magnitude of the CH/ $\pi$  interaction in the gas phase: experimental and theoretical determination of the accurate interaction energy in benzene-methane. *J. Phys. Chem. A* **110**, 4397–4404 (2006).
- Meot-Ner, M. & Deakne, C. A. Unconventional ionic hydrogen-bonds. 1. CH<sup>+</sup>...X. Complexes of quaternary ions with n- and  $\pi$ -donors. *J. Am. Chem. Soc.* **107**, 469–474 (1985).
- Osborne, M. R. & Phillips, D. H. Preparation of a methylated DNA standard, and its stability on storage. *Chem. Res. Toxicol.* **13**, 257–261 (2000).
- Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).
- Besler, B. H., Merz, K. M. & Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **11**, 431–439 (1990).
- Ramstein, J. & Lavery, R. Energetic coupling between DNA bending and base pair opening. *Proc. Natl Acad. Sci. USA* **85**, 7231–7235 (1988).
- Adhikary, S., Cato, M. C., McGary, K. L., Rokas, A. & Eichman, B. F. Non-productive DNA damage binding by DNA glycosylase-like protein Mag2 from *Schizosaccharomyces pombe*. *DNA Repair (Amst.)* **12**, 196–204 (2013).
- Dalhous, B. et al. Sculpting of DNA at abasic sites by DNA glycosylase homolog Mag2. *Structure* **21**, 154–166 (2013).

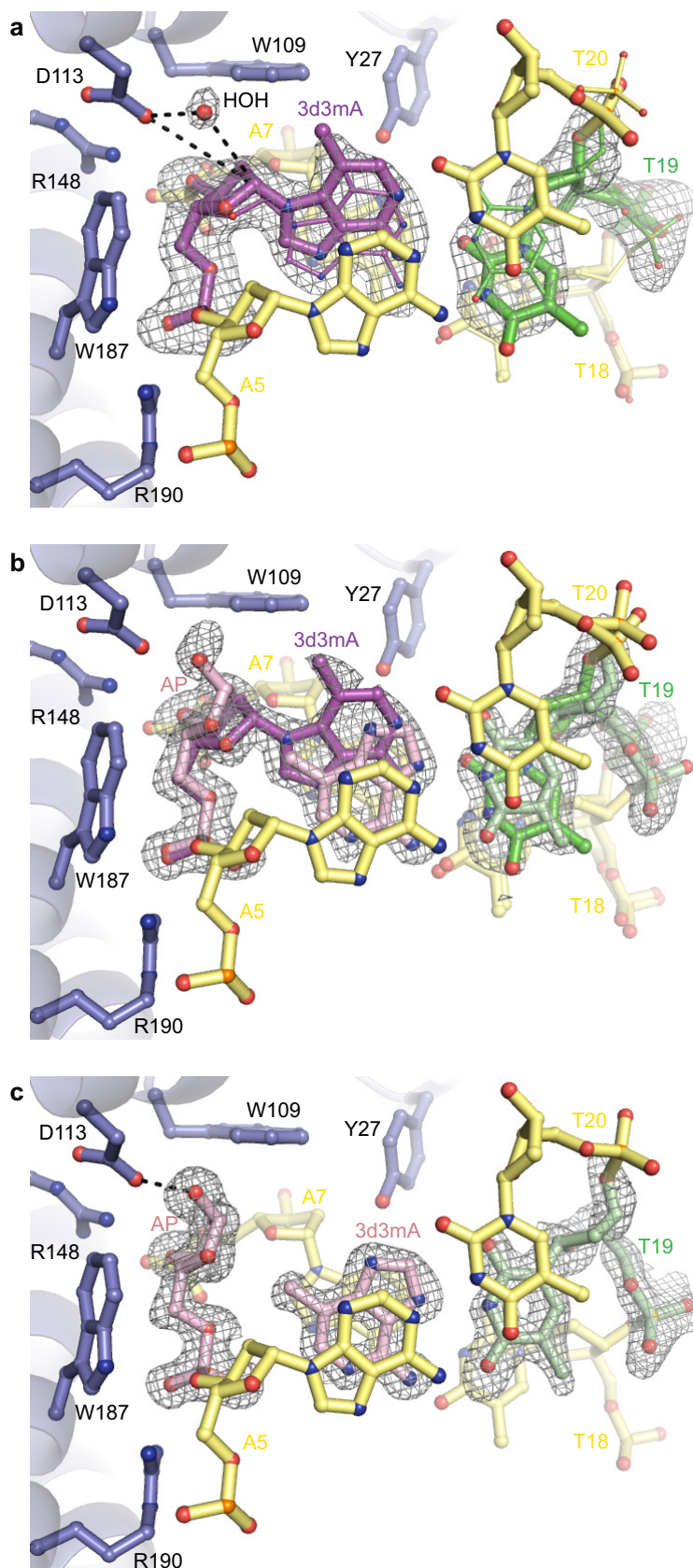




**Extended Data Figure 1 | Remodelled DNA structures.** DNA binding by AlkD bends the helical axis by  $30^\circ$  and widens the minor groove by 4 Å. Remodelling of this type reduces the energetic barrier to base pair opening but does not force base flipping<sup>54–56</sup>. In the AlkD–3d3mA–DNA complex, this distortion creates an equilibrium between Watson–Crick and sheared conformations of the 3d3mA•T base pair. In the sheared conformation, the 3d3mA nucleobase is displaced by 4 Å into the widened minor groove but remains partially stacked in the DNA duplex. **a**, Sheared 3d3mA•T base

pair in the AlkD–DNA complex. **b**, Watson–Crick 3d3mA•T base pair in the AlkD–DNA complex. **c**, Simulated Watson–Crick A•T base pair in the AlkD–DNA complex. Adenine was generated by removing the methyl substituent from 3d3mA and changing carbon to nitrogen at the 3-position. Computational optimization was not performed. **d**, Watson–Crick A•T base pair in B-form DNA (PDB accession 1BNA). The asterisks in **a** and **b** indicate the position of the methyl substituent at C3. All structures are depicted as stereo images.

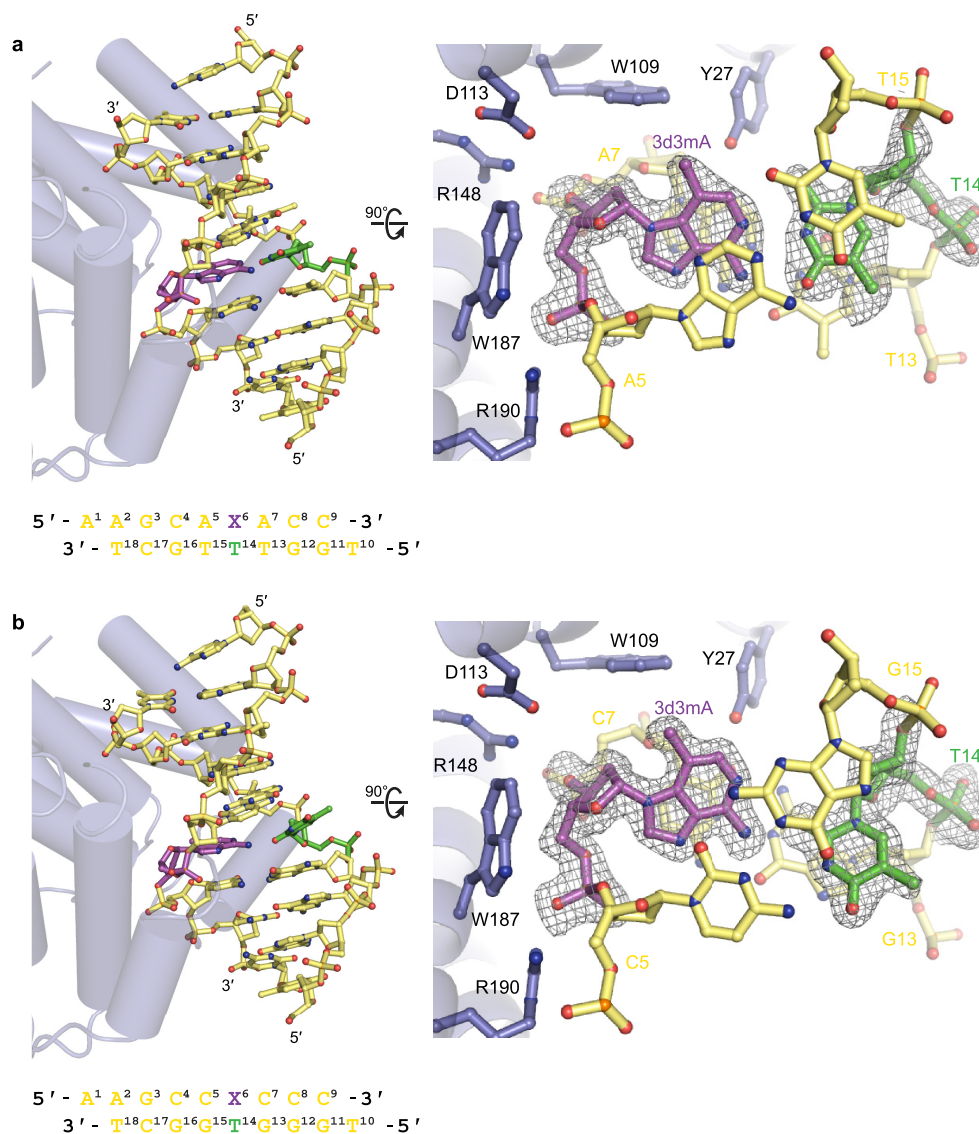




**Extended Data Figure 2 | 3d3mA substrate and product structures.** The AlkD-3d3mA-DNA complex was crystallized under moderately acidic (pH 5.7) conditions, causing a small fraction (~1%) of 3d3mA to become protonated and activated for enzymatic excision. Reaction progress was monitored by flash-freezing crystals at various times after preparing the protein-DNA complex and refining the fractional occupancies of substrate and product in the structures. **a**, Structure after 4 h containing only

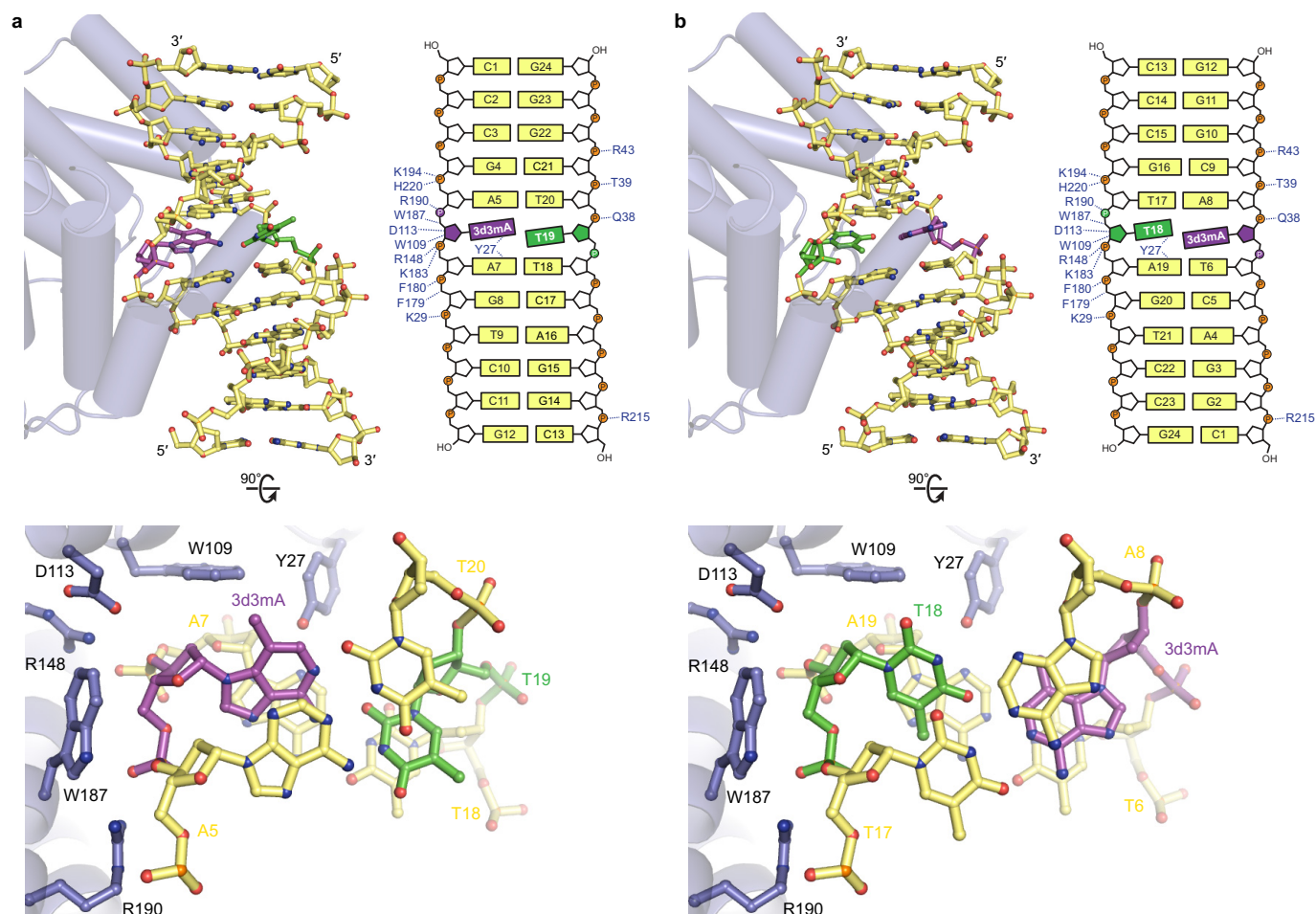
3d3mA-DNA substrate. The 3d3mA•T base pair is present as a mixture of Watson-Crick (thin bonds) and sheared (thick bonds) conformations.

**b**, Structure after 48 h containing a mixture of 3d3mA-DNA substrate together with AP-DNA product and 3d3mA nucleobase. **c**, Structure after 360 h containing only AP-DNA product and 3d3mA nucleobase. All panels show annealed omit electron density contoured to  $2.5\sigma$ .



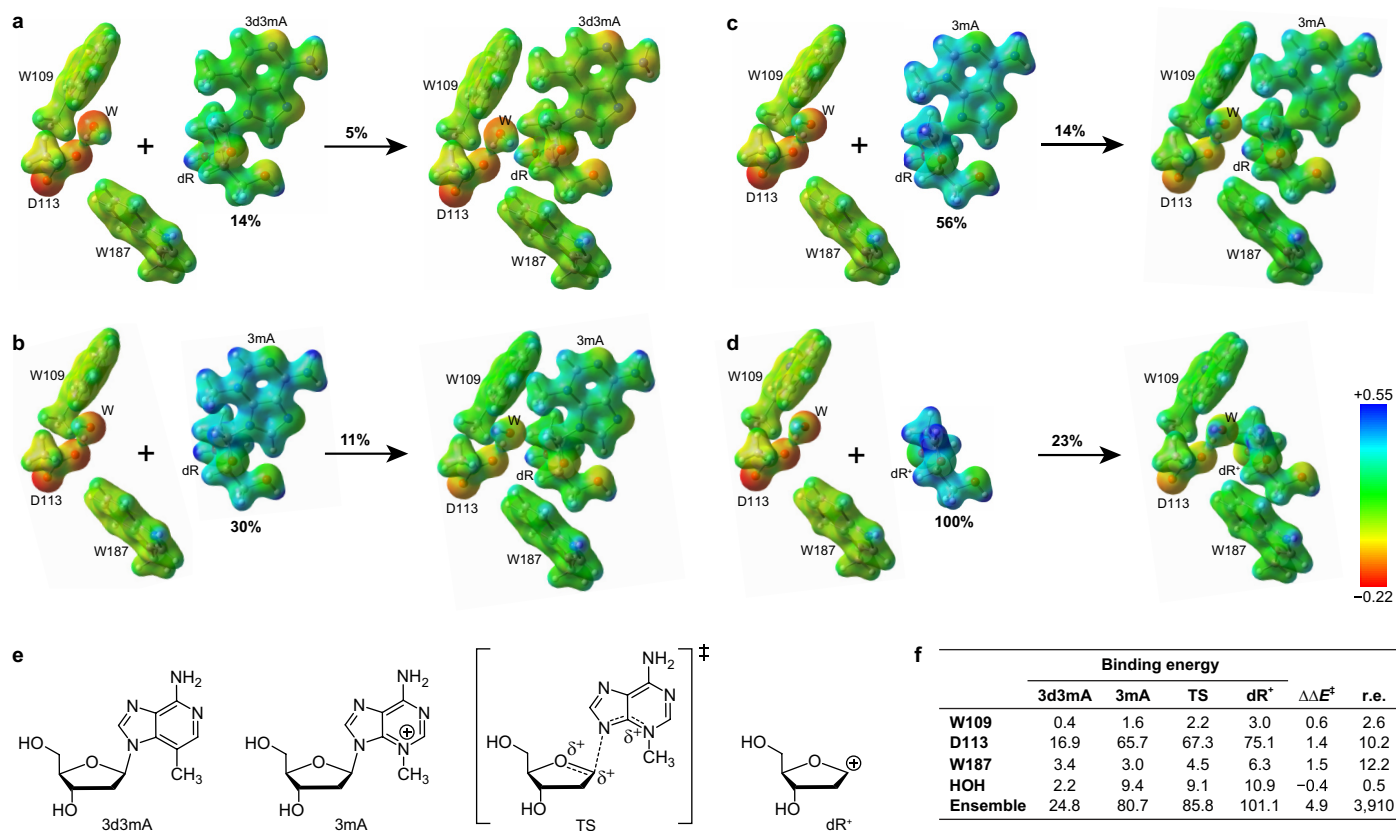
**Extended Data Figure 3 | Additional substrate structures.** AlkD was crystallized with two alternate 9-mer 3d3mA-DNA constructs to verify that crystal packing contacts were not significantly influencing the conformation of the DNA. As in the catalytic 12-mer substrate complex, 3d3mA remained stacked in the duplex in both 9-mer structures. However, unlike in the 12-mer complex, a mixture of Watson–Crick and sheared conformations was not observed. Instead, the 3d3mA•T base pair in each construct

is either entirely in the Watson–Crick conformation or entirely in the sheared conformation. Formation of the AP product was not observed in either 9-mer construct, probably as a consequence of the neutral (pH 7.0) crystallization conditions, under which the fraction (<0.1%) of 3d3mA activated for depurination would be drastically reduced. **a**, Watson–Crick conformation. **b**, Sheared conformation. Both panels show annealed omit electron density contoured to  $2.5\sigma$ .



**Extended Data Figure 4 | Comparison of catalytic and non-catalytic AlkD-3d3mA-DNA complexes.** AlkD binds 3d3mA-DNA in two orientations that differ only in the position of the 3d3mA•T base pair about its dyad axis. Either 3d3mA (catalytic complex) or the opposing thymine (non-catalytic complex) resides against the protein surface. Both orientations use a common set of interactions that induce bending of the helical axis and widening of the minor groove, causing shearing of the 3d3mA•T base pair. In either complex, the nucleotide adjacent to the protein surface shows the same degree of displacement into the minor groove. The asymmetric position of the 3d3mA nucleotide in the duplex allows

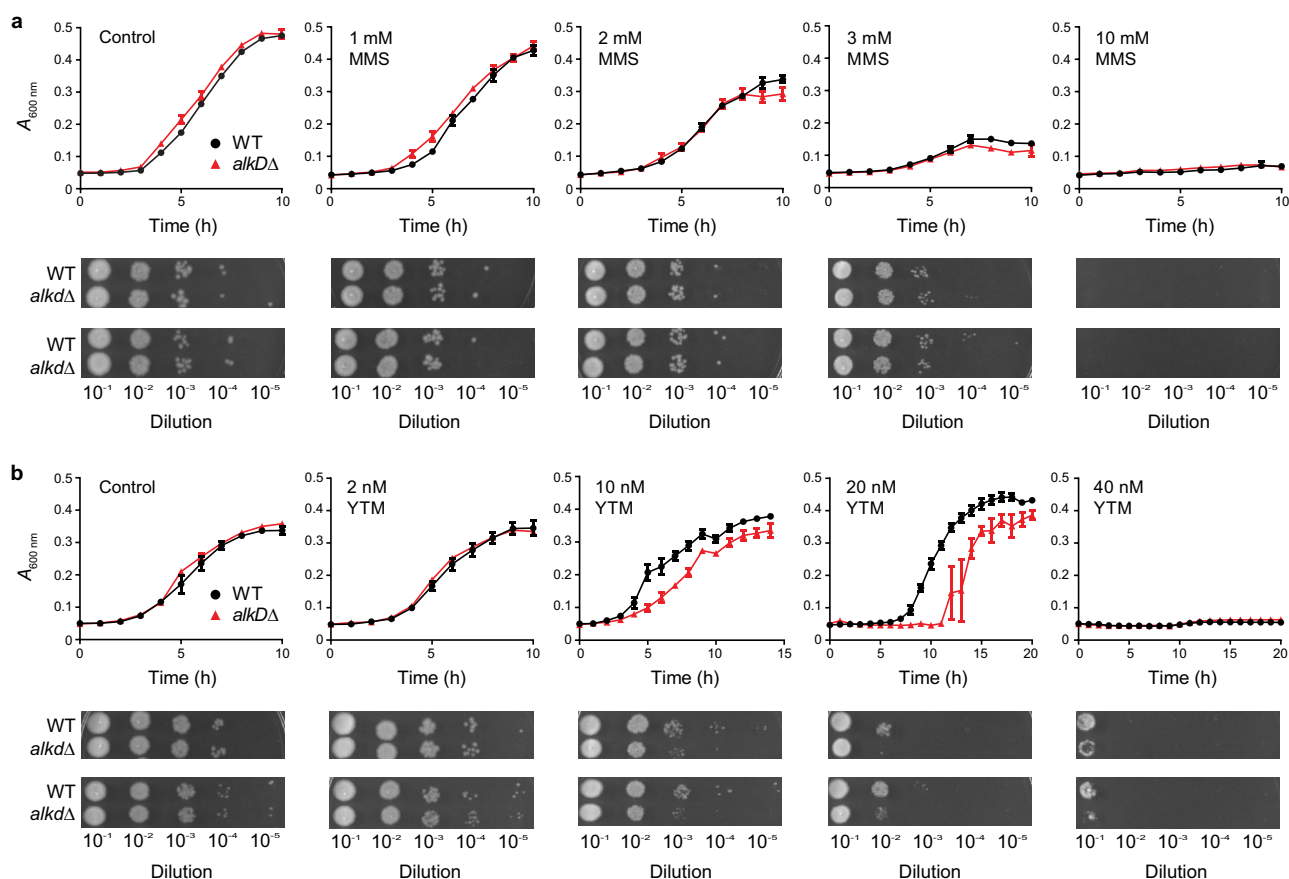
crystal packing interactions to select for a single orientation from a likely mixture of orientations in solution. Different DNA constructs were used to exploit these packing interactions and crystallize complexes in each binding orientation separately. Positive charge on the deoxyribose of 3mA would produce stronger electrostatic interactions with Trp109, Asp113 and Trp187 that would likely favour the catalytic binding orientation as well as a sheared conformation of a 3mA•T base pair. **a**, Catalytic orientation. The Watson-Crick conformer of the 3d3mA•T base pair was omitted for clarity. **b**, Non-catalytic orientation (PDB accession 3JX7).



**Extended Data Figure 5 | Electrostatic potential surfaces and computed binding energies.** Charge transfer from the catalytic ensemble (Trp109, Asp113, Trp187 and Asp113-associated water) to the modified nucleosides stabilizes the AlkD–DNA complexes. **a**, 3d3mA substrate. **b**, 3mA substrate. **c**, Transition state (TS) approximation. **d**, dR<sup>+</sup> intermediate. Positive charge located on the deoxyribose moiety of the unbound nucleosides relative to that on the unbound dR<sup>+</sup> intermediate is indicated below the lesions.

Charge transferred from the catalytic ensemble to the lesions upon complex formation is indicated above the arrows. Electrostatic potentials were scaled to  $-0.22$ – $0.55$  atomic units on an isodensity surface of  $0.05$  electrons bohr<sup>-3</sup>. **e**, Modified nucleosides shown in **a**–**d**. **f**, Computed binding energies (kcal mol<sup>-1</sup>), differential bond elongation energies ( $\Delta\Delta E^\ddagger$ , kcal mol<sup>-1</sup>), and corresponding rate enhancements (r.e.).





**Extended Data Figure 6 | Growth curves and spot assays.** Wild-type (black) and *alkD*-knockout (red) strains of *B. anthracis* were treated with varying concentrations of MMS and YTM. Control experiments contained no drug. Deletion of AlkD caused no observable phenotype with MMS

but resulted in increased sensitivity to YTM. **a**, MMS treatment. **b**, YTM treatment. Error bars represent the s.e.m. from three replicate growth curves. Spot assays were performed in duplicate.

Extended Data Table 1 | X-ray data collection and refinement statistics

	3d3mA•T-DNA 100% substrate 4 hours	3d3mA•T-DNA AP•T-DNA/3d3mA 71% substrate/29% product 24 hours	3d3mA•T-DNA AP•T-DNA/3d3mA 51% substrate/49% product 48 hours	3d3mA•T-DNA AP•T-DNA/3d3mA 33% substrate/67% product 72 hours
<b>Data collection</b>				
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions				
$a, b, c$ (Å)	38.19, 93.65, 48.08	38.25, 93.56, 47.98	38.34, 93.40, 48.12	38.39, 93.28, 48.08
$\alpha, \beta, \gamma$ (°)	90.00, 113.05, 90.00	90.00, 112.74, 90.00	90.00, 112.85, 90.00	90.00, 112.72, 90.00
Resolution (Å)	50.00–1.97 (2.04–1.97)*	50.00–1.88 (1.95–1.88)	50.00–1.57 (1.60–1.57)	50.00–1.54 (1.57–1.54)
$R_{\text{merge}}$	0.068 (0.521)	0.084 (0.590)	0.063 (0.422)	0.058 (0.184)
Avg. $I/\sigma I$	18.4 (2.9)	18.4 (3.2)	21.3 (3.1)	34.0 (10.3)
Completeness (%)	99.8 (99.5)	99.9 (99.9)	99.9 (98.9)	99.4 (93.9)
Redundancy	4.1 (4.0)	5.3 (4.5)	4.2 (3.7)	7.5 (6.8)
<b>Refinement</b>				
Resolution (Å)	40.01–1.97 (2.06–1.97)	35.28–1.87 (1.94–1.87)	35.34–1.57 (1.61–1.57)	35.06–1.54 (1.58–1.54)
No. reflections	21,918 (2,701)	25,801 (2,791)	43,431 (3,010)	45,847 (3,225)
$R_{\text{work}}$	0.156 (0.217)	0.153 (0.216)	0.158 (0.208)	0.153 (0.169)
$R_{\text{free}}^{\dagger}$	0.189 (0.305)	0.196 (0.267)	0.189 (0.238)	0.183 (0.211)
No. atoms				
Protein	1,951	1,951	1,951	1,951
DNA	487	487	487	488
Water	177	207	356	413
Avg. $B$ -factors <sup>‡</sup> (Å <sup>2</sup> )				
Protein	28.5	27.7	15.7	18.4
DNA	50.1	46.2	34.0	38.0
Water	39.0	38.6	31.8	35.2
R.m.s. deviations				
Bond lengths (Å)	0.007	0.007	0.006	0.006
Bond angles (°)	0.990	1.021	1.052	1.043
Ramachandran distribution (%)				
Favored	97.8	97.8	97.4	97.8
Allowed	2.2	2.2	2.6	2.2
Disallowed	0.0	0.0	0.0	0.0
PDB accession code	5CL3	5CL4	5CL5	5CL6

\*Statistics for the highest resolution shell are shown in parentheses.

<sup>†</sup> $R_{\text{free}}$  was determined from the 5% of reflections excluded from refinement.<sup>‡</sup>Equivalent isotropic  $B$ -factors were calculated in conjunction with TLS-derived anisotropic  $B$ -factors.

Extended Data Table 2 | X-ray data collection and refinement statistics

	3d3mA•T-DNA AP•T-DNA/3d3mA 18% substrate/82% product 96 hours	AP•T-DNA/3d3mA 100% product 144 hours	AP•T-DNA/3d3mA 100% product 240 hours	AP•T-DNA/3d3mA 100% product 360 hours
<b>Data collection</b>				
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions				
$a, b, c$ (Å)	38.39, 93.15, 48.20	38.50, 93.45, 48.77	37.90, 92.87, 48.92	38.43, 93.08, 48.05
$\alpha, \beta, \gamma$ (°)	90.00, 112.81, 90.00	90.00, 114.19, 90.00	90.00, 113.77, 90.00	90.00, 112.61, 90.00
Resolution (Å)	50.00–1.44 (1.46–1.44)*	50.00–1.38 (1.40–1.38)	50.00–1.54 (1.57–1.54)	50.00–1.54 (1.60–1.54)
$R_{\text{merge}}$	0.069 (0.509)	0.047 (0.318)	0.050 (0.523)	0.049 (0.138)
Avg. $I/\sigma I$	23.3 (3.5)	31.9 (5.1)	32.0 (3.0)	41.1 (16.6)
Completeness (%)	96.4 (93.8)	99.5 (96.1)	99.5 (94.3)	96.9 (92.3)
Redundancy	6.4 (6.1)	7.3 (5.2)	6.2 (4.8)	7.7 (7.4)
<b>Refinement</b>				
Resolution (Å)	35.40–1.44 (1.47–1.44)	35.13–1.38 (1.41–1.38)	35.03–1.54 (1.58–1.54)	35.05–1.54 (1.58–1.54)
No. reflections	54,660 (3,702)	64,110 (4,402)	45,352 (3,149)	44,651 (3,037)
$R_{\text{work}}$	0.128 (0.150)	0.133 (0.160)	0.160 (0.227)	0.148 (0.147)
$R_{\text{free}}^{\dagger}$	0.166 (0.216)	0.158 (0.231)	0.186 (0.268)	0.169 (0.194)
No. atoms				
Protein	1,951	1,951	1,951	1,951
DNA	488	488	488	488
Water	363	405	248	400
Avg. $B$ -factors <sup>‡</sup> (Å <sup>2</sup> )				
Protein	17.0	18.2	26.6	16.2
DNA	38.1	39.7	53.6	35.1
Water	33.2	33.9	41.1	33.5
R.m.s. deviations				
Bond lengths (Å)	0.008	0.008	0.006	0.006
Bond angles (°)	1.219	1.157	0.972	1.041
Ramachandran distribution (%)				
Favored	98.3	98.3	97.8	97.9
Allowed	1.7	1.7	2.2	2.1
Disallowed	0.0	0.0	0.0	0.0
PDB accession code	5CL7	5CL8	5CL9	5CLA

\*Statistics for the highest resolution shell are shown in parentheses.

<sup>†</sup> $R_{\text{free}}$  was determined from the 5% of reflections excluded from refinement.<sup>‡</sup>Equivalent isotropic  $B$ -factors were calculated in conjunction with TLS-derived and individual anisotropic  $B$ -factors.

Extended Data Table 3 | X-ray data collection and refinement statistics

	3d3mA•T-DNA 9-mer A	3d3mA•T-DNA 9-mer B	1aR•T-DNA/3mA	THF•T-DNA/3mA
<b>Data collection</b>				
Space group	<i>C</i> 2	<i>C</i> 2	<i>P</i> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub>
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	127.44, 55.41, 47.97	127.40, 55.45, 47.56	38.20, 92.83, 48.31	38.49, 93.29, 48.16
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 105.70, 90.00	90.00, 106.36, 90.00	90.00, 113.71, 90.00	90.00, 112.84, 90.00
Resolution (Å)	50.00–1.77 (1.83–1.77)*	50.00–1.73 (1.79–1.73)	50.00–1.54 (1.57–1.54)	50.00–1.73 (1.79–1.73)
<i>R</i> <sub>merge</sub>	0.088 (0.414)	0.072 (0.362)	0.057 (0.255)	0.067 (0.443)
Avg. <i>I</i> / $\sigma$ <i>I</i>	15.0 (2.7)	21.3 (5.1)	78.6 (7.0)	23.2 (3.4)
Completeness (%)	96.7 (80.0)	100.0 (99.9)	97.2 (92.2)	99.9 (99.6)
Redundancy	4.6 (4.2)	5.3 (4.8)	7.1 (6.4)	4.9 (4.6)
<b>Refinement</b>				
Resolution (Å)	46.20–1.77 (1.82–1.77)	45.65–1.73 (1.78–1.73)	35.00–1.54 (1.58–1.54)	40.09–1.73 (1.78–1.73)
No. reflections	30,638 (2,200)	33,412 (2,770)	44,194 (3,013)	32,539 (2,620)
<i>R</i> <sub>work</sub>	0.163 (0.228)	0.154 (0.196)	0.149 (0.153)	0.155 (0.211)
<i>R</i> <sub>free</sub> <sup>†</sup>	0.203 (0.277)	0.186 (0.252)	0.179 (0.179)	0.192 (0.288)
No. atoms				
Protein	1,898	1,898	1,951	1,951
DNA	364	364	487	487
Water	253	284	415	271
Avg. <i>B</i> -factors <sup>‡</sup> (Å <sup>2</sup> )				
Protein	24.2	23.6	14.6	21.2
DNA	38.3	32.2	33.1	39.5
Water	34.0	34.3	31.9	35.5
R.m.s. deviations				
Bond lengths (Å)	0.007	0.007	0.006	0.007
Bond angles (°)	1.030	1.074	1.049	1.009
Ramachandran distribution (%)				
Favored	97.7	97.3	97.9	97.8
Allowed	2.3	2.7	2.1	2.2
Disallowed	0.0	0.0	0.0	0.0
PDB accession code	5CLB	5CLC	5CLD	5CLE

\*Statistics for the highest resolution shell are shown in parentheses.

<sup>†</sup>*R*<sub>free</sub> was determined from the 5% of reflections excluded from refinement.<sup>‡</sup>Equivalent isotropic *B*-factors were calculated in conjunction with TLS-derived anisotropic *B*-factors.



# Structure of a eukaryotic SWEET transporter in a homotrimeric complex

Yuyong Tao<sup>1\*</sup>, Lily S. Cheung<sup>2\*</sup>, Shuo Li<sup>1,3</sup>, Joon-Seob Eom<sup>2</sup>, Li-Qing Chen<sup>2</sup>, Yan Xu<sup>1</sup>, Kay Perry<sup>4</sup>, Wolf B. Frommer<sup>2</sup> & Liang Feng<sup>1</sup>

Eukaryotes rely on efficient distribution of energy and carbon skeletons between organs in the form of sugars. Glucose in animals and sucrose in plants serve as the dominant distribution forms. Cellular sugar uptake and release require vesicular and/or plasma membrane transport proteins. Humans and plants use proteins from three superfamilies for sugar translocation: the major facilitator superfamily (MFS), the sodium solute symporter family (SSF; only in the animal kingdom), and SWEETs<sup>1–5</sup>. SWEETs carry mono- and disaccharides<sup>6</sup> across vacuolar or plasma membranes. Plant SWEETs play key roles in sugar translocation between compartments, cells, and organs, notably in nectar secretion<sup>7</sup>, phloem loading for long distance translocation<sup>8</sup>, pollen nutrition<sup>9</sup>, and seed filling<sup>10</sup>. Plant SWEETs cause pathogen susceptibility possibly by sugar leakage from infected cells<sup>3,11,12</sup>. The vacuolar *Arabidopsis thaliana* AtSWEET2 sequesters sugars in root vacuoles; loss-of-function mutants show increased susceptibility to *Pythium* infection<sup>13</sup>. Here we show that its orthologue, the vacuolar glucose transporter OsSWEET2b from rice (*Oryza sativa*), consists of an asymmetrical pair of triple-helix bundles, connected by an inversion linker transmembrane helix (TM4) to create the translocation pathway. Structural and biochemical analyses show OsSWEET2b in an apparent inward (cytosolic) open state forming homomeric trimers. TM4 tightly interacts with the first triple-helix bundle within a protomer and mediates key contacts among protomers. Structure-guided mutagenesis of the close paralogue SWEET1 from *Arabidopsis* identified key residues in substrate translocation and protomer crosstalk. Insights into the structure–function relationship of SWEETs are valuable for understanding the transport mechanism of eukaryotic SWEETs and may be useful for engineering sugar flux.

SWEET transporters are unique in that they have seven predicted transmembrane (TM) domains with two internal triple-helix bundles (THBs). Prokaryotic SemiSWEETs, which also transport mono- or disaccharides<sup>4,14–16</sup>, contain only a single THB<sup>4,14</sup>. SWEETs likely arose by gene duplication of the prokaryotic unit in concert with the insertion of an inversion linker helix<sup>4</sup>. Structural analyses demonstrated that two SemiSWEETs dimerize in a parallel configuration to create a translocation pore<sup>14–16</sup>. Based on the SemiSWEET structure, a single SWEET unit with its two THBs was proposed to form the translocation path<sup>14</sup>. However, single SWEETs may not function autonomously, because they can form functionally important oligomers<sup>4</sup>. Low sequence conservation and the unpredictable effect of the inversion linker TM domain on the structure of eukaryotic SWEETs limit our mechanistic understanding of the transport process.

To elucidate the molecular transport mechanism of eukaryotic SWEETs and determine the relevance of oligomerization for sugar transport, we carried out a systematic analysis of eukaryotic SWEETs to identify candidates suitable for structural and biochemical studies. OsSWEET2b from rice *O. sativa* featured desirable biochemical

properties. Phylogenetically, OsSWEET2b belongs to clade I (refs 3, 6, 7), which includes the founding family member, *A. thaliana* SWEET1 (AtSWEET1), a plasma membrane hexose transporter<sup>3</sup> (Extended Data Fig. 1), and the closest relative of OsSWEET2b, AtSWEET2, a transporter involved in vacuolar retention of sugars and pathogen resistance<sup>13</sup>. Similar to AtSWEET2, OsSWEET2b–EGFP localized to vacuolar membranes when expressed in yeast (Fig. 1a). Consistent with the intracellular localization, OsSWEET2b did not complement the glucose uptake deficiency of a yeast plasma membrane hexose transporter mutant (Extended Data Fig. 2).

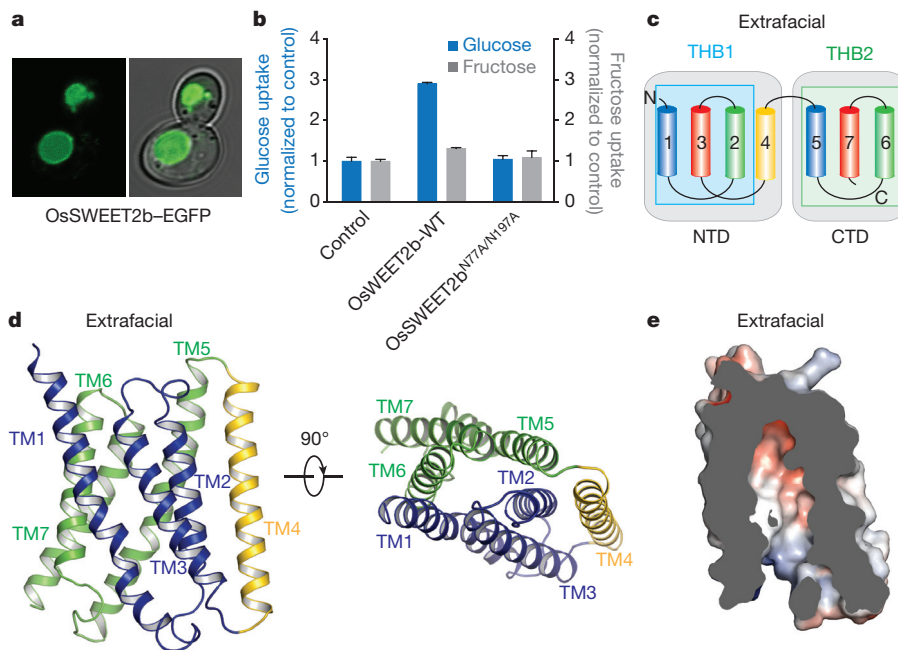
To directly test activity, OsSWEET2b was purified and reconstituted into lipid vesicles. Vesicles containing OsSWEET2b showed elevated glucose transport activity compared to control vesicles; activity was abolished when using OsSWEET2b mutations known to block SWEET transport activity<sup>14</sup> (Fig. 1b). By contrast, fructose uptake was marginal and close to the detection limit (Fig. 1b). A chimera between the amino (N)-terminal (first four TMs) domain of OsSWEET2b and the carboxy (C)-terminal (last three TMs) domain of OsSWEET1a, a close clade I homologue of OsSWEET2b, resulted in robust glucose transport activity (Extended Data Fig. 3). Together, these results implicate OsSWEET2b in glucose transport, similar to other clade I and II SWEET transporters. As AtSWEET1 is closely related to OsSWEET2b (Extended Data Fig. 4), can be measured in a variety of heterologous transport assays, and has the greatest quantity of available functional data, it was used for structure–function studies directed by the OsSWEET2b structure, as described below.

We crystallized and solved OsSWEET2b structures in two forms at 3.1 Å ( $P_{21}$ ) and 3.7 Å ( $P_{21}2_12_1$ ) resolution (Extended Data Table 1). Experimental phases were determined by single-wavelength anomalous dispersion (SAD) on selenomethionine-containing crystals (Extended Data Fig. 5a). For one asymmetric unit of either crystal form, three OsSWEET2b molecules are related by a non-crystallographic three-fold axis perpendicular to the membrane plane, forming a trimeric assembly (Fig. 3a, b and Extended Data Fig. 5b, c). Individual OsSWEET2b copies superimposed with a root mean square deviation (r.m.s.d.) of less than 0.8 Å, indicating nearly identical overall structures.

Each OsSWEET2b protomer is composed of seven TM helices. The N- and C-terminal three TM helices form THBs with a characteristic 1–3–2 helix arrangement and a similar fold to SemiSWEET THBs<sup>14–16</sup> (Fig. 1c, d). Comparison of OsSWEET2b with SemiSWEET revealed that six TMs in an OsSWEET2b protomer resemble the functional dimer of SemiSWEET where the N-terminal THB1 and C-terminal THB2 enclose a putative transport route (Fig. 2a). This indicates that a SWEET protomer is sufficient to form the pore. TM4, designated as an inversion linker, spans the membrane and orients THB2 parallel to THB1, creating a similar configuration as in homodimeric SemiSWEETs<sup>14</sup>. TM4 packs intimately against THB1 to constitute the N-terminal domain of SWEET, but barely contacts THB2 within the same protomer (Fig. 1c, d). This feature explains previously puzzling

<sup>1</sup>Department of Molecular and Cellular Physiology, 279 Campus Drive, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>2</sup>Carnegie Institution for Science, Department of Plant Biology, 260 Panama Street, Stanford, California 94305, USA. <sup>3</sup>Center of Growth, Metabolism and Aging, Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610014, China. <sup>4</sup>NE-CAT and Dep. of Chemistry and Chemical Biology, Cornell University, Building 436E, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, Illinois 60439, USA.

\*These authors contributed equally to this work.



**Figure 1 | Localization and structure of OsSWEET2b.** **a**, OsSWEET2b-EGFP localizes to the vacuolar membrane in yeast. **b**, Transport activity of OsSWEET2b was examined by a liposome uptake assay. Glucose uptake is shown in blue. Liposomes with OsSWEET2b had elevated glucose uptake compared with control liposomes without protein. Mutation of two putative pore-lining asparagine residues abolished transport activities ( $\pm$  s.e.m.,  $n = 4$ ). Fructose uptake is shown in grey ( $\pm$  s.e.m.,  $n = 3$ ). **c**, Membrane topology

diagram of OsSWEET2b. TM4 and THB1 constitute the N-terminal domain (NTD), while THB2 forms the C-terminal domain (CTD). **d**, Ribbon representation of the OsSWEET2b protomer. THB1 is coloured in blue, THB2 in green, and TM4 in yellow. **e**, Slab view of OsSWEET2b in an inward (cytoplasmic) open conformation. The surface is coloured according to electrostatic potential.

observations that co-expression of AtSWEET1 split as THB1+TM4 and THB2 reconstituted transport activity, while THB1 and TM4+THB2 did not<sup>4</sup>. To accommodate TM4, the intrafacial (cytosolic) half of THB1 bends away from where TM4 binds, causing significant structural divergence of THB1 and THB2 (Extended Data Fig. 6a). Together, the membrane topology of SWEET shows different connectivity and spatial transmembrane arrangement when compared to other known heptahelical membrane proteins, in particular G-protein-coupled receptors or PnuC vitamin transporters<sup>17–19</sup> (Extended Data Fig. 7).

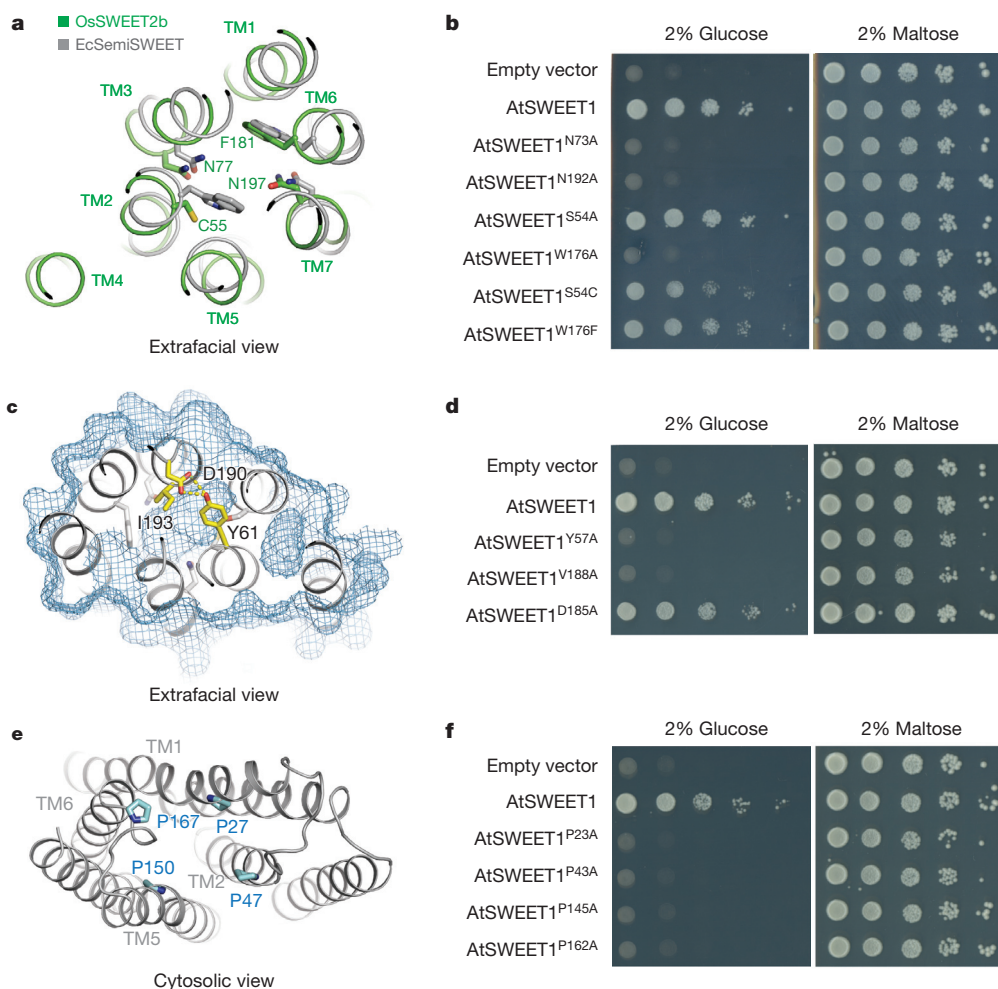
Phosphorylation of the C-terminus of a plasma membrane SWEET supports the idea that it faces the cytosol<sup>20</sup>. By analogy, we predict that the C-terminus of the vacuolar OsSWEET2b faces the cytosol (intrafacial side), while the extrafacial side locates in the vacuolar lumen (Fig. 1c). Between the N- and C-terminal domains of each protomer, a large solvent-accessible cavity extends from the cytosolic side to the top of the molecule where it is sealed from the vacuolar side by an extrafacial gate (Fig. 1e). The structure thus appears to represent an inward (cytosolic) open conformation. A major constituent of the extrafacial gate is Tyr61 on THB1 (Fig. 2c and Extended Data Fig. 8c) that hydrogen-bonds with Asp190 and interacts with Gln132 (Fig. 2c). The Tyr–Asp pair is highly conserved (Extended Data Fig. 4). The side chain of Ile193 also forms part of the luminal gate (Fig. 2c). Replacing residues equivalent to Tyr61 and Ile193 in AtSWEET1 (Tyr57 and Val188) abolished transport activity, confirming their importance (Fig. 2d). Consistent with roles in transport, the majority of conserved residues of plant SWEETs line the cavity, with two highly conserved clusters near the predicted substrate-binding pocket and the putative intrafacial gate (Extended Data Fig. 8a, b). A series of hydrophobic residues line the transport pathway, potentially lowering resistance for sugar passage (Fig. 1e).

Structural comparison of OsSWEET2b with the *Escherichia coli* EcSemiSWEET that was crystallized in an inward-open conformation<sup>16</sup> reveals common themes and distinct features. THB1 of OsSWEET2b

differs significantly from the protomer of EcSemiSWEET (r.m.s.d. = 4.1 Å), while THB2 superimposes better (r.m.s.d. = 2.4 Å; Extended Data Fig. 6b). OsSWEET2b residues involved in possible substrate binding or extrafacial gate formation are positioned close to those in EcSemiSWEET<sup>16</sup> (Fig. 2a). This indicates that heptahelical eukaryotic SWEETs and tri-helical prokaryotic SemiSWEETs share common sugar-binding pockets and transport routes. Consistent with this, the extrafacial half of OsSWEET2b overlays better with EcSemiSWEET than the intrafacial half.

Both the transport pathway and the putative substrate-binding pocket of OsSWEET2b are highly asymmetrical, contrasting with the symmetry of the ancestral SemiSWEET homodimers. In SemiSWEETs, pairs of tryptophans and asparagines are invariant, forming the binding pocket and facilitating transport<sup>14–16</sup>. In OsSWEET2b, two highly conserved asparagine residues (Asn77 and Asn197) surround the putative substrate-binding pocket at equivalent positions (Fig. 2a and Extended Data Fig. 8c). Both are essential for AtSWEET1 activity<sup>14</sup> (Fig. 2b). In contrast, no tryptophan pairs were found along the cavity of OsSWEET2b. Instead, Cys58 on THB1 and Phe181 on THB2 occupy equivalent positions in OsSWEET2b, while a Ser54 and Trp176 are the equivalents in AtSWEET1 (Fig. 2a and Extended Data Fig. 4). Replacing Trp176 of AtSWEET1 with alanine abolished activity, whereas alanine- or cysteine-substitution of Ser54 had a minimal effect on activity (Fig. 2b). Furthermore, substituting Trp176 in AtSWEET1 with phenylalanine had little effect on transport activity (Fig. 2b). Together these results indicate that one aromatic residue in THB2, but not THB1, is essential for transport, highlighting the asymmetrical configuration of SWEET, in which the two THBs diverged to create an asymmetric binding pocket in eukaryotic SWEETs.

Near the intrafacial side, TM1, TM2, TM5, and TM6 are the main constituents of the transport route (Fig. 2e). Notably, in each of these four TMs, a proline that either terminates the helix or induces a kink faces the transport route at a similar level in the membrane (Fig. 2e). Substituting any of the four prolines caused loss of AtSWEET1 activity



**Figure 2 | Key elements of the transport pathway.** **a**, Comparison of the substrate-binding pocket of OsSWEET2b (green) to that of EcSemiSWEET (grey). Key residues in the binding pocket are shown as sticks. Labels are for OsSWEET2b. **b**, AtSWEET1 complements the defective growth in glucose of EBY4000, a hexose transport-deficient yeast strain. W176A, N73A, and N192A mutations in AtSWEET1 abolished glucose-dependent growth. Growth on maltose containing medium was not affected by the different mutants. **c**, Extrafacial (extracellular) gate. Residues that cap the aqueous cavity from the extrafacial side form the extrafacial gate with side chains shown as sticks in yellow. Residues in the substrate-binding pocket are coloured in grey. The

surface of the protein is shown as a pale blue mesh. **d**, Effect of mutations in the extrafacial gate of AtSWEET1. Alanine substitution of V188, but not the conserved D185, in the extrafacial gate of AtSWEET1 abolished growth in glucose. Alanine substitution of Y57 leads to a growth defect on glucose, possibly due to mistargeting<sup>4</sup>. **e**, Intrafacial hinge points. The four proline residues (shown as sticks) lining the transport pathway are near the same level in the membrane. **f**, The intrafacial (cytosolic) gate is required for sugar transport. Mutations in the conserved prolines that are important for the intrafacial gate of AtSWEET1 (P23A, P43A, P145A and P162A) eliminate the growth of the EBY4000 yeast strain in synthetic medium supplemented with glucose.

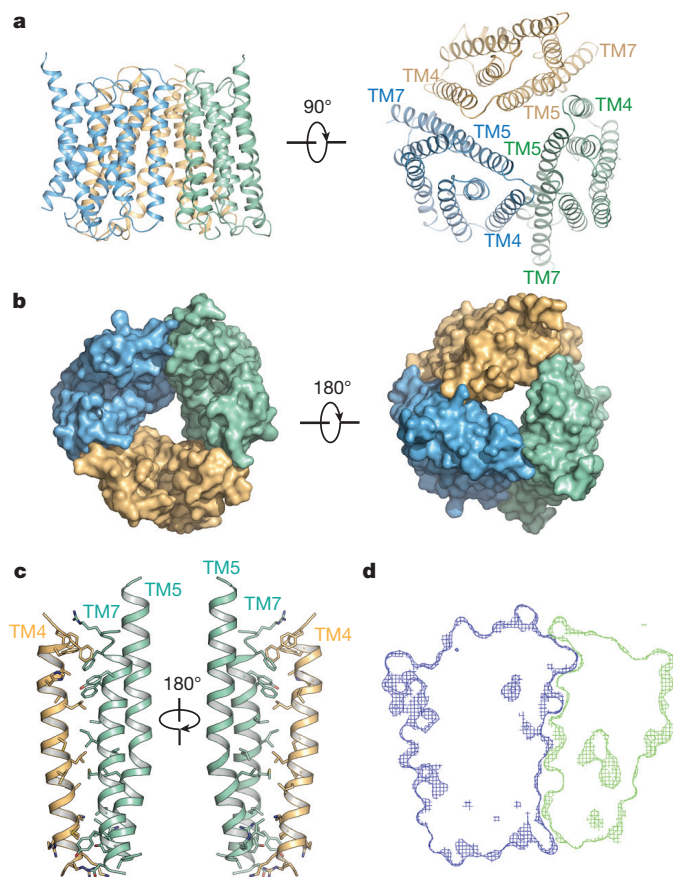
(Fig. 2f and Extended Data Fig. 9). These observations, together with the strict conservation across all *Arabidopsis* SWEETs, support important roles of the proline tetrad in the transport mechanism. The proline ring might enable concerted structural rearrangements associated with transitions between conformational states, possibly as hinges for gating the transport pathway. Interestingly, a highly conserved PQ motif is found at the equivalent position on TM1 in SemiSWEETs<sup>14–16</sup>. A glutamine serves as a key component of the cytosolic gate through cross-protomer hydrogen bonds. The absence of a glutamine next to the proline in eukaryotic SWEETs suggests that equivalent cross-protomer hydrogen bonds are not required to form the cytosolic gate, which is likely made possible by covalent linkage through TM4. Residues immediately above and below the proline ring in the pore are highly conserved, and substitution with alanine reduced activity (Extended Data Fig. 8d), indicating that neighbouring residues may also contribute to or affect the intrafacial gate.

Our structural and functional observations delineate the transport route to within an individual SWEET protomer. Crystal structures reveal three OsSWEET2b units forming a trimer (Fig. 3a, b).

Interactions between the protomers are primarily mediated by TM4 of one protomer with TM7 as well as the C-terminal half of TM5 of the neighbouring protomer (Fig. 3a, c). Several lines of evidence support a physiological relevance of trimerization. First, the relative position of the three protomers is compatible with a trimer in the membrane plane (Fig. 3a). Second, the interface between protomers is extensive, burying  $\sim 1700 \text{ \AA}^2$  between two protomers with a shape complementarity index score of 0.69, similar as in typical antibody–antigen pairs<sup>21</sup> (Fig. 3d). Third, an identical trimeric assembly was observed in two crystal forms with different packing modes, indicating that trimer formation is independent of specific crystal packing (Extended Data Fig. 5b, c).

To validate the trimeric assembly, we performed biochemical characterizations. Co-immunoprecipitation experiments suggest that both OsSWEET2b and AtSWEET1 form homo-oligomers (Fig. 4a). Furthermore, glutaraldehyde crosslinked OsSWEET2b into trimers in a concentration-dependent manner (Fig. 4b), supporting trimer formation in solution. To dissect whether the observed trimer of OsSWEET2b represents its quaternary structure in the membrane,

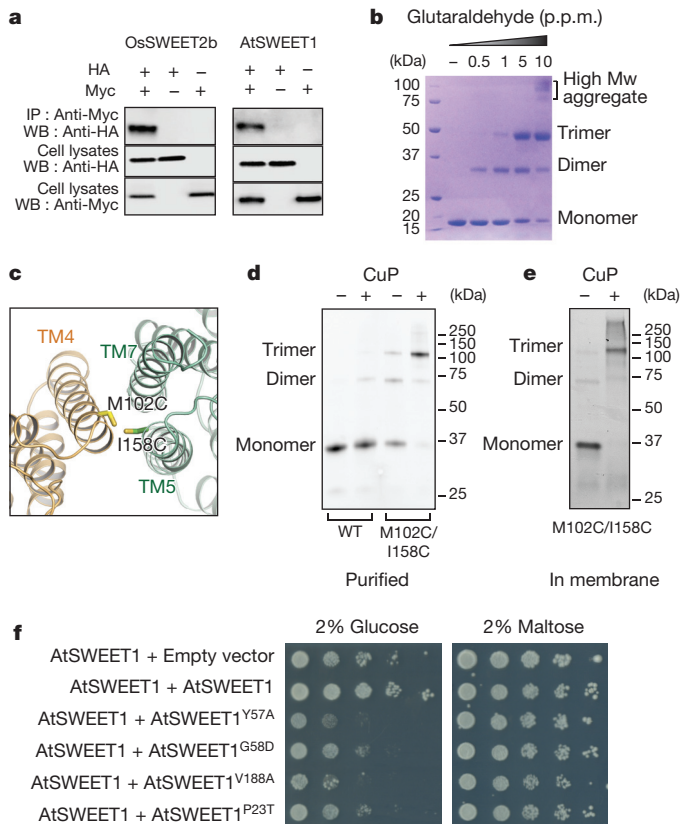




**Figure 3 | Structure of the OsSWEET2b trimer.** **a**, Two orthogonal views of the OsSWEET2b trimer in ribbon representation. **b**, Surface representations of OsSWEET2b viewed from the intrafacial (cytosolic, left) and extrafacial (luminal, right) side. **c**, Close-up view of the trimer interface. TM4 from one protomer and TM5 and TM7 from the neighbouring protomer are shown as ribbon representation. Side chains of residues participating in the interaction are shown as sticks. **d**, Cross-section view of the trimer interface. The two interacting protomers are presented as blue and green surface meshes.

we carried out site-directed disulfide-bridge crosslinking studies. The double cysteine mutant with a pair of cysteines at the trimer interface (Fig. 4c) crosslinked via disulfide bridges under oxidizing conditions in both detergent micelles and in the isolated membrane, as shown by the newly formed crosslink product at the expected trimer position on SDS-PAGE (Fig. 4d, e). Together, these data support trimeric assembly as observed in crystal structures to represent the *in vivo* quaternary structure.

The trimeric assembly of OsSWEET2b provides a structural basis for explaining observations from split-GFP and split-ubiquitin experiments that SWEETs homo- or hetero-oligomerize<sup>4</sup>. Mutagenesis identified dominant negative mutations of AtSWEET1 that inhibit the activity of co-expressed wild-type proteins; a finding originally considered to indicate transport route formation from multiple subunits<sup>4</sup>. The trimeric structure of OsSWEET2b offers a new explanation: each protomer encompasses a transport route, though protomers might be coupled allosterically. Two previously identified dominant-negative mutations, Y57A and its neighbouring mutant G58D, were expected to disrupt the extrafacial gate or interfere with its conformational transitions given that Y57 is a major component of this gate. Though mistargeting of the Y57 mutant complicates the interpretation, the G58 mutant is correctly localized to the plasma membrane, and has a similar dominant-negative effect. This suggests that interference with the proper function of the extrafacial gate in one protomer may hinder activity of other protomers, consistent with allosteric coupling. To further assess the role of the extrafacial gate, we examined



**Figure 4 | Trimer formation by OsSWEET2b.** **a**, Co-immunoprecipitation (Co-IP) of OsSWEET2b and AtSWEET1. Left, c-Myc-tagged OsSWEET2b pulls down HA-tagged OsSWEET2b; right, c-Myc-tagged AtSWEET1 pulls down HA-tagged AtSWEET1. **b**, Crosslink of purified OsSWEET2b in detergent micelle. Increasing amounts of glutaraldehyde (0, 0.5, 1, 5, 10 p.p.m) were incubated with the purified OsSWEET2b. The samples were analysed by SDS-PAGE. Mw, molecular weight. **c**, The design of cysteine pair mutations at the interface of the trimer. M102 and I158 were mutated to cysteines (side chains shown as sticks). **d**, Cysteine-directed crosslinking of OsSWEET2b in detergent solution. GFP-tagged OsSWEET2b with M102C/I158C mutations was purified and oxidized in air or with copper phenanthroline (CuP). Proteins separated on SDS-PAGE were imaged by in-gel fluorescence of GFP. **e**, Cysteine-directed crosslinking of OsSWEET2b in membrane. Cell membranes containing the M102C/I158C mutant were untreated or treated with CuP. OsSWEET2b-GFP was visualized by in-gel fluorescence. **f**, Mutations in the residues at the gates of AtSWEET1 have a dominant negative effect on transport. EBY4000 yeast co-expressing wild-type AtSWEET1 and either an empty vector control or wild-type AtSWEET1 were able to grow on medium supplemented with glucose as the sole carbon source. Co-expression with an AtSWEET1(Y57A) or G58D mutant prevents growth on glucose. Similarly, co-expression with mutants that abolish transport and that are localized in the intrafacial (P23T) or extrafacial (V188A) gates prevent growth on glucose. As a control, growth on maltose was not affected.

V188, the other main contributor to this gate. As shown above, V188A mutation abolishes transport activity in AtSWEET1 (Fig. 2d). When co-expressed with wild-type AtSWEET1, the V188A mutant almost completely inhibited activity of the wild-type form (Fig. 4f). In addition, we confirmed that the P23T mutation causes dominant negative inhibition of a co-expressed wild-type transporter<sup>4</sup>, possibly due to its effect on the cytosolic gate. These gate mutations might impair conformational state transitions, as observed in other transporters<sup>22,23</sup>. Together, these results show that gate mutations cause dominant negative effects and support that structure and function are coupled among SWEET protomers. As a variety of SWEETs form homo- or heterooligomers<sup>4</sup>, this coupling may be more generally applicable, expanding the properties and regulatory states of SWEET sugar transporters.



The proposed structural coupling of protomers may affect conformational states and cause functional coupling, which bears similarities to cooperativity found in other classes of transporters. For instance, human GLUT1 forms tetramers in which each protomer transports glucose, however protomers interact cooperatively such that two subunits exist in outward-open and two as inward-open states<sup>24</sup>. Similarly, protomers of triple-barrelled ammonium transporters (AMTs) are allosterically coupled in a way that transport activity is feedback-inhibited by its own substrate via phosphorylation<sup>25,26</sup>.

Membrane transporters with internal symmetry are widespread: notable examples include MFS, LeuT-fold, AMT, CLC, and ABC transporters<sup>27,28</sup>. These transport proteins are thought to have evolved from 'half-transporters' containing a single repeat via gene duplication and subsequent fusion<sup>28</sup>. The availability of the structures of OsSWEET2b and several SemiSWEET transporters provides a structural framework to understand the link between half-transporters and full-transporters with internal repeats. In eukaryotic SWEETs, covalent fusion of two THBs through an inversion TM creates an asymmetrical translocation pathway, greatly enriching their structural features compared to the simpler symmetrical SemiSWEETs. Covalent fusion in eukaryotic SWEETs thus expands their functional capacity to accommodate diverse substrates, enhance substrate specificity or fine-tune activity.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May; accepted 10 August 2015.

Published online 19 October 2015.

- Wright, E. M. Glucose transport families SLC5 and SLC50. *Mol. Aspects Med.* **34**, 183–196 (2013).
- Cura, A. J. & Carruthers, A. Role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism, and homeostasis. *Compr. Physiol.* **2**, 863–914 (2012).
- Chen, L. Q. *et al.* Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527–532 (2010).
- Xuan, Y. H. *et al.* Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *Proc. Natl Acad. Sci. USA* **110**, E3685–E3694 (2013).
- Chen, L. Q., Cheung, L. S., Feng, L., Tanner, W. & Frommer, W. B. Transport of sugars. *Annu. Rev. Biochem.* **84**, 865–894 (2015).
- Eom, J. S. *et al.* SWEETs, transporters for intracellular and intercellular sugar translocation. *Curr. Opin. Plant Biol.* **25**, 53–62 (2015).
- Lin, I. W. *et al.* Nectar secretion requires sucrose phosphate synthases and the sugar transporter SWEET9. *Nature* **508**, 546–549 (2014).
- Chen, L. Q. *et al.* Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* **335**, 207–211 (2012).
- Sun, M. X., Huang, X. Y., Yang, J., Guan, Y. F. & Yang, Z. N. *Arabidopsis* RPG1 is important for primexine deposition and functions redundantly with RPG2 for plant fertility at the late reproductive stage. *Plant Reprod.* **26**, 83–91 (2013).
- Chen, L. Q. *et al.* A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the *Arabidopsis* embryo. *Plant Cell* **27**, 607–619 (2015).
- Zhou, J. *et al.* Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *Plant J.* **82**, 632–643 (2015).
- Cohn, M. *et al.* *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. *Mol. Plant Microbe Interact.* **27**, 1186–1198 (2014).
- Chen, H.-Y. *et al.* Vacuolar SWEET2 transporter reduces glucose efflux to limit pathogen propagation in roots. *Plant J.* **83**, 1046–1058 (2015).
- Xu, Y. *et al.* Structures of bacterial homologues of SWEET transporters in two distinct conformations. *Nature* **515**, 448–452 (2014).
- Wang, J. *et al.* Crystal structure of a bacterial homologue of SWEET transporters. *Cell Res.* **24**, 1486–1489 (2014).
- Lee, Y., Nishizawa, T., Yamashita, K., Ishitani, R. & Nureki, O. Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter. *Nat. Commun.* **6**, 6112 (2015).
- Yee, D. C. *et al.* The transporter-opsin-G protein-coupled receptor (TOG) superfamily. *FEBS J.* **280**, 5780–5800 (2013).
- Jaehme, M., Guskov, A. & Slotboom, D. J. Crystal structure of the vitamin B3 transporter PnuC, a full-length SWEET homologue. *Nature Struct. Mol. Biol.* **21**, 1013–1015 (2014).
- Jaehme, M., Guskov, A. & Slotboom, D. J. The twisted relation between Pnu and SWEET transporters. *Trends Biochem. Sci.* **40**, 183–188 (2015).
- Niittylä, T., Fuglsang, A. T., Palmgren, M. G., Frommer, W. B. & Schulze, W. X. Temporal analysis of sucrose-induced phosphorylation changes in plasma membrane proteins of *Arabidopsis*. *Mol. Cell. Proteomics* **6**, 1711–1726 (2007).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Kazmier, K. *et al.* Conformational dynamics of ligand-dependent alternating access in LeuT. *Nature Struct. Mol. Biol.* **21**, 472–479 (2014).
- Krishnamurthy, H. & Gouaux, E. X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* **481**, 469–474 (2012).
- De Zutter, J. K., Levine, K. B., Deng, D. & Carruthers, A. Sequence determinants of GLUT1 oligomerization: analysis by homology-scanning mutagenesis. *J. Biol. Chem.* **288**, 20734–20744 (2013).
- Yuan, L. *et al.* Allosteric regulation of transport activity by heterotrimerization of *Arabidopsis* ammonium transporter complexes *in vivo*. *Plant Cell* **25**, 974–984 (2013).
- Langar, V. *et al.* Feedback inhibition of ammonium uptake by a phospho-dependent allosteric mechanism in *Arabidopsis*. *Plant Cell* **21**, 3610–3622 (2009).
- Forrest, L. R., Kramer, R. & Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta* **1807**, 167–188 (2011).
- Keller, R., Ziegler, C. & Schneider, D. When two turn into one: evolution of membrane transporters from half modules. *Biol. Chem.* **395**, 1379–1388 (2014).

**Acknowledgements** We thank the staff at Beamline 23ID-B and 23ID-D (APS, Argonne National Laboratory) and at Beamline 5.0.2 (Advanced Light Source) for assistance at the synchrotron. This work was made possible by support from Stanford University, the Harold and Leila Y. Mathers Charitable Foundation and the Alfred P. Sloan Foundation to L.F. and the Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences at the US Department of Energy (DOE DE-FG02-04ER15542) to W.B.F. The functional characterization of OsSWEET2b was supported by grants from the National Science Foundation (IOS-1258018) to W.B.F. L.F. is a Klingenstein-Simons Fellow. L.S.C. was supported by the National Science Foundation Postdoctoral Research Fellowship in Biology (1401855). S.L. was supported by the National Natural Science Foundation of China (31300618). Part of this work was conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team, supported by a grant from the National Institute of General Medical Sciences (NIH, P41 GM103403). Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) and the Office of Science by Argonne National Laboratory, was supported by US DOE (DE-AC02-06CH11357).

**Author Contributions** W.B.F. and L.F. conceived and designed experiments. Y.T. performed expression, purification, crystallization, data collection, crystallography, and biochemical experiments, S.L. performed liposome uptake experiments, Y.X. contributed to initial characterization and biochemical characterizations, L.S.C., L.-Q.C., and J.-S.E. performed functional experiments, K.P. assisted crystallography, L.F. contributed to initial characterization, data collection and crystallography. Y.T., L.C., L.-Q.C., J.-S.E., S.L., W.B.F., and L.F. analysed the data. L.F. and W.B.F. wrote the manuscript.

**Author Information** The crystallographic coordinates and structural factors are deposited into Protein Data Bank with accession numbers 5CTG (3.1 Å) and 5CTH (3.7 Å). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.F. (liangf@stanford.edu).

## METHODS

**Protein expression and purification.** The *Oryza sativa* OsSWEET2b coding sequence was cloned into a modified pPIC3-C vector (Invitrogen Life Technologies) with C-terminal 3C protease cleavage site, EGFP and His-tag. To obtain diffracting crystals, the intrinsically disordered and poorly conserved 15 residues at the C-terminal after TM7 were removed (Extended Data Fig. 4). Transformed *Pichia pastoris* were grown until  $A_{600\text{nm}}$  ( $\text{OD}_{600}$ ) = 10 and induced at 27 °C. Membrane proteins of lysed cells were extracted by 3% dodecyl- $\beta$ -D-maltoside (DDM, Anatrace) in the presence of protease inhibitor cocktail at 4 °C for 2 h. The insoluble fractions were removed by centrifugation at 16,000 r.p.m for 45 min. The supernatant was incubated with cobalt resin at 4 °C for 2 h. The resin was washed with buffer containing 0.3% *n*-decyl- $\beta$ -D-maltoside (DM, Anatrace) and 20 mM imidazole. The target protein was cleaved off from the tag by incubating with 3C protease. For crystallization experiments, the protein surface was modified by reductive methylation<sup>29</sup>. The protein was further purified by gel filtration (Superdex 200, GE Healthcare) in buffer containing 10 mM Tris-HCl (pH 8.0), 150 mM NaCl and 0.5% *n*-nonyl- $\beta$ -D-glucopyranoside (NG, Anatrace). The peak fraction was collected and concentrated to 10 mg ml<sup>-1</sup> for crystallization trials. For nonspecific crosslinking, the protein was extracted and purified in DDM. After the cobalt affinity purification and 3C protease cleavage, the protein was directly subject to further purification by gel filtration in 10 mM HEPES-NaOH (pH 8.0), 150 mM NaCl and 0.05% DDM.

To produce selenomethionine-labelled OsSWEET2b protein, Sf9 insect cells (*Spodoptera frugiperda*) in methionine-free medium were infected by baculovirus (Bac-to-Bac system, Invitrogen) containing OsSWEET2b construct and were supplemented with 100 mg l<sup>-1</sup> selenomethionine after 12 h. Cells were collected after 48 h before the membrane fractions were isolated by ultracentrifugation. The protein was purified in the same manner as the unlabelled protein.

**Crystallization and structural determination.** The crystals were grown using the hanging drop method by mixing protein with an equal volume of crystallization solution at 22 °C. Plate-like crystals (Form I and SeMet crystal) were grown in 0.1 M MES (pH 6.5), 30% PEG400, 10 mM MnCl<sub>2</sub>, 5% ethanol. Cubic crystals (Form II crystal) were grown in 0.1 M MES (pH 6.0), 27% PEG400. The Form I crystals were directly flash-frozen in liquid nitrogen. The Form II crystals were cryo-protected and dehydrated by gradually increasing PEG400 concentration to 44% and then flash-frozen in liquid nitrogen.

The diffraction data were collected at 24ID-C/E and 23ID-B/D of the Advanced Photon Source and 5.0.2 of Advanced Light Source. Native data set was processed using HKL2000 (ref. 30). The selenomethionine anomalous data were collected from multiple crystals and integrated using XDS<sup>31</sup> as implemented in RAPD (<https://github.com/RAPD/RAPD>). The eight most correlated partial data sets from three crystals, as determined by the Pearson's coefficient between common Bragg reflections, were combined with POINTLESS and scaled with AIMLESS<sup>32</sup>. The Se sites and initial phases were determined by SHELX C/D/E<sup>33</sup> and improved by Phaser-EP<sup>34</sup>. The map was improved by Resolve<sup>35</sup> and was of sufficient quality to identify and trace transmembrane helices with the aid of Se positions. The model built into the map was directly refined against the native data set in P<sub>2</sub><sub>1</sub> space group (Form I). The resulting model was subject to iterative refinement with NCS restraints in Phenix and manual rebuilding in Coot<sup>36,37</sup>. The final model has 98.8% residues in the favoured region with the rest in the allowed region from the Ramachandran plot. For P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> data set (Form II), the molecular replacement solution was found using the refined model from Form I by Phaser<sup>38</sup>. The model was first refined by DEN and then iteratively by refine.phenix<sup>37,39</sup>. In the final model, 95.9% residues are in the favoured region and the rest are in the allowed region. Structure quality was assessed by MolProbity and figures were prepared based on higher resolution P<sub>2</sub><sub>1</sub> form using PyMOL<sup>40,41</sup>.

**Crosslinking.** For nonspecific crosslinking, purified protein in DDM (around 1 mg ml<sup>-1</sup> in 15  $\mu$ l) was incubated with various concentrations of glutaraldehyde (0, 0.0005%, 0.001%, 0.005%, 0.01%) at room temperature for 8 h. The reaction was quenched by addition of 50 mM Tris-Cl, pH 8.0. The sample was mixed with equal volume of SDS loading buffer, separated on 10% SDS-PAGE gels and stained with Coomassie blue. For site-directed cysteine crosslinking of protein in detergent micelles, the GFP-tagged protein was purified by cobalt affinity column in 0.05% DDM and incubated with 0.3 mM CuSO<sub>4</sub> and 0.9 mM o-phenanthroline for 2 h at 4 °C. The reaction was quenched by 50 mM EDTA. For site-directed cysteine crosslinking in membrane, the membrane fraction was prepared and incubated with 0.3 mM CuSO<sub>4</sub> and 0.9 mM o-phenanthroline for 2 h at 4 °C. The reaction was stopped by the addition of 50 mM EDTA. All samples were mixed with equal volume of SDS loading buffer and separated on 10% SDS-PAGE gels. The gels were imaged using a Bio-Rad imaging system with excitation and emission wavelengths of 488 nm and 522 nm, respectively.

**Liposome-based uptake assays.** Liposomes were prepared according to a published protocol<sup>42</sup>. Purified OsSWEET2b was reconstituted into *E. coli* polar lipid

(Avanti) with a protein to lipid ratio of 1:80 in KPM6.5 buffer (50 mM potassium phosphate, 2 mM MgSO<sub>4</sub>, pH 6.5). The proteoliposomes were extruded through 400 nm membrane and harvested by ultracentrifugation at 100,000g for 1 h. The proteoliposomes were washed twice with ice-cold KPM6.5 buffer and resuspended in the same buffer to a final concentration of 100 mg ml<sup>-1</sup> immediately before the assay. For each uptake assay, 2  $\mu$ l of proteoliposomes were diluted into 100  $\mu$ l of assay buffer containing 0.2  $\mu$ Ci of D-[<sup>14</sup>C]glucose in the presence of 5 mM cold glucose. The reaction mixtures were incubated for 2 h or 3 h at room temperature, and then rapidly filtered through 0.22  $\mu$ m cellulose acetate filter membranes (Millipore). The filter membranes were immediately washed with 2 ml ice-cold reaction buffer without any substrate and taken for liquid scintillation counting. The uptake of fructose was performed similarly. All the experiments were repeated at least three times independently.

**Yeast complementation assay and fluorescence imaging.** The transport activity of AtSWEET1 and its mutants was tested by complementing the growth defect of the hexose uptake mutant yeast strain EBY4000 [*hxt1-17A::loxP gal2A::loxP stl1A::loxP agt1A::loxP ydl247wA::loxP yjr160cA::loxP*]<sup>43</sup>. The coding sequence for AtSWEET1 and mutants were cloned in the vectors pDRf1-GW, p112A1NE-GW and pDRf1-GFP GW vectors for yeast expression. Expression of the transgene in p112A1NE is under the control of the ADH1 promoter, while overexpression in the pDRf1 vector is regulated by the strong PMA promoter. Yeast was transformed using the lithium acetate/SS carrier DNA/PEG method<sup>44</sup>. Transformants were selected on synthetic medium (0.17% YNB and 0.5% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>) supplemented with 2% maltose and auxotrophic requirements. Cells were grown in liquid synthetic medium overnight and then diluted to an  $A_{600\text{nm}}$  ( $\text{OD}_{600}$ ) of ~0.2 in water before plating. Tenfold serial dilutions were spotted on medium containing auxotrophic requirements and 2% glucose as the sole carbon source or 2% maltose as a control. Plates were incubated at 30 °C for 4 days, and imaged using a CanoScan scanner (Canon). Representative samples from three separate transformations are shown.

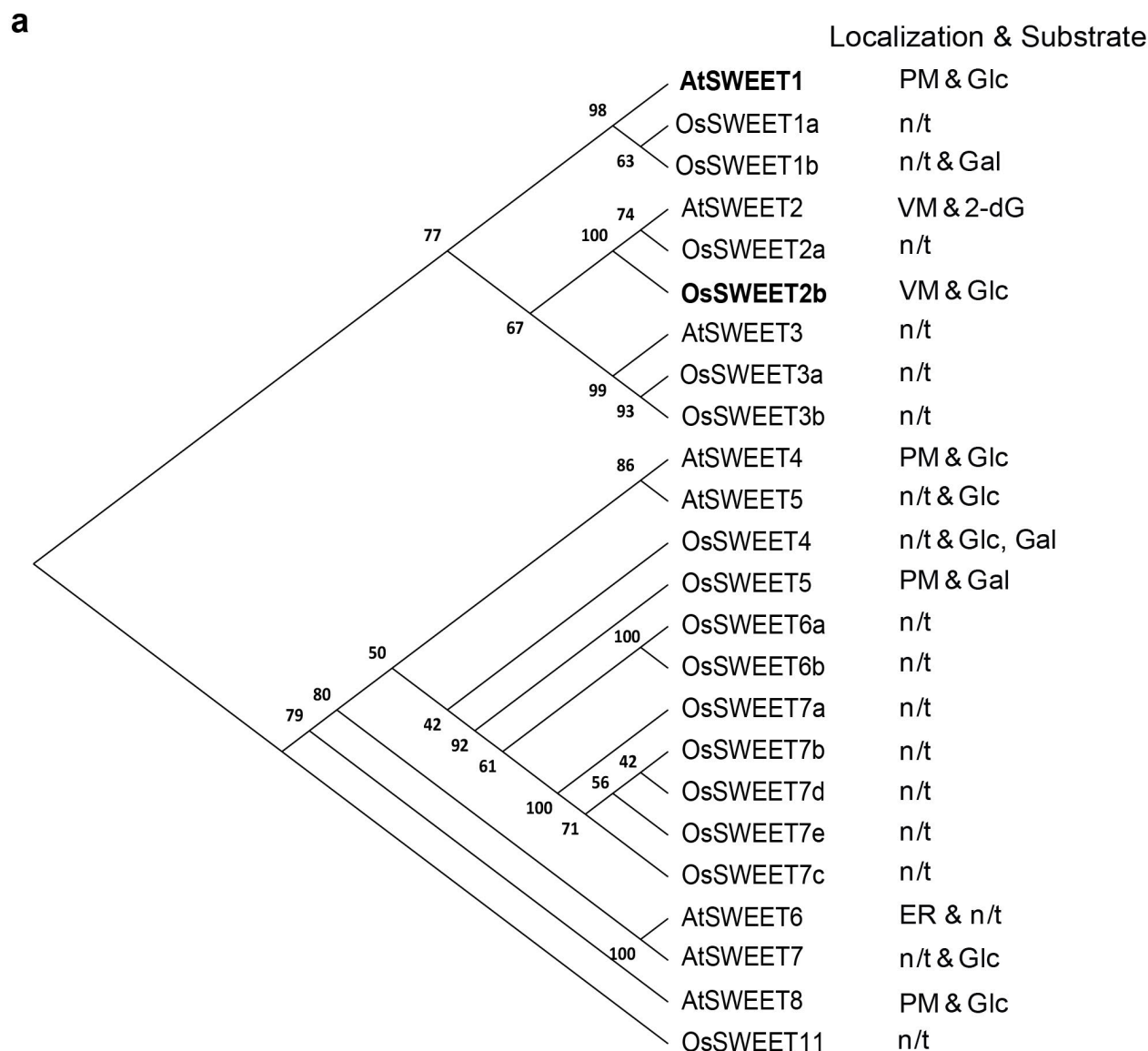
Confocal imaging of EGFP tagged AtSWEET1 and mutants was performed on a Leica TCS SP5 microscope with a HCX PL APO lambda blue 63 $\times$ , 1.40 NA, oil UV objective, 488 nm excitation and 492–600 nm emission. Images were autocontrasted and overlaid with Fiji (<http://fiji.sc>). Cells were grown overnight in liquid synthetic medium supplemented with 2% maltose and auxotrophic requirements, and resuspended in water before imaging.

**Co-immunoprecipitation.** OsSWEET2b and AtSWEET1 coding sequences without stop codon were Gateway cloned (Life Technologies) into the yeast expression vectors pDR-GW- Trp1-4 $\times$ Myc and pDR-GW- Leu2-3 $\times$ HA carrying either the TRP1 or the LEU2 auxotrophic markers. These vectors were derived from pDRf1GW<sup>45</sup> by replacing the URA3 marker by a TRP1 (pDR-GW-TRP1) or a LEU2 marker (pDR-GW-LEU2) to allow for co-expression and by adding 4 copies of the c-Myc tag or 3 copies of the HA-tag. Tagged proteins were expressed in the protease-deficient *Saccharomyces cerevisiae* strain BJ5465 [MATa *ura3-52 trp1 leu2A1 his3A200 pep4::HIS3 prb1A1.6R can1 GAL*] (ATCC). Yeast coexpressing either cMyc- and HA-tagged OsSWEET2b, or cMyc- and HA-tagged AtSWEET1, were selected on synthetic medium (0.17% YNB and 0.5% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>) supplemented with 2% glucose without leucine and tryptophan. Cells were grown in liquid synthetic medium until  $A_{600\text{nm}}$  ( $\text{OD}_{600}$ ) 0.4–0.6, and harvested by centrifugation at 4,000g for 10 min. Cell lysates were extracted in buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM EDTA, 10% glycerol, 0.1% NP-40, 1% Triton X-100, 0.1% SDS and protease inhibitor cocktail (Roche). Cell lysates were incubated for 4 h with 1  $\mu$ g anti-Myc (Santa Cruz Biotechnology; catalogue number sc-789) antibody in the presence of Protein A/G agarose beads (Santa Cruz Biotechnology). The resulting complexes were washed four times with a washing buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.1% NP-40, 1% Triton X-100 and protease inhibitor cocktail, followed by acid elution using 0.2 M glycine pH 2.6 on ice for 10 min. Eluates were neutralized by adding 1 M Tris pH 7.5. Anti-HA antibody (1:1,000, Roche; catalogue number 11 867 423 001) and secondary anti-rat-HRP (1:10,000, Pierce catalogue number 31470) and anti-rabbit-HRP (1:10,000, Bio-Rad catalogue number 1706515) were used for immunoblotting.

**Transporter activity assay using FRET glucose sensor in HEK293T cells.** Procedures for cell culture, transfection, image acquisition and FRET sensor analysis have been described previously<sup>46</sup>. Briefly, HEK293T cells (ATCC CRL-11268, without further authentication or test of mycoplasma contamination) were cotransfected with a plasmid carrying the glucose sensor FLI<sup>12</sup>Pglu700u86 only, and either OsSWEET1a or a chimeric plasmid OsSWEET2b<sup>1–127</sup>OsSWEET1a<sup>124–273</sup> carrying OsSWEET2b (1–127 amino acids) fused with OsSWEET1a (124–273 amino acids), which was obtained by overlapping PCR. One day after transfection, imaging was carried out at the following settings: exposure time 300 ms, gain 3, binning 2, and time interval 10 s. Data were normalized to the initial ratio.

**Data reporting.** No statistical methods were used to predetermine sample size.

29. Rypniewski, W. R., Holden, H. M. & Rayment, I. Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8-Å resolution. *Biochemistry* **32**, 9851–9858 (1993).
30. Otwinowski, Z. & Minor, W. in *Methods Enzymol.* Vol. 276 (ed. Carter, C. W. Jr) 307–326 (Academic Press, 1997).
31. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133–144 (2010).
32. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D* **67**, 282–292 (2011).
33. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
34. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
35. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
37. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
38. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32–41 (2007).
39. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
40. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
41. DeLano, W. L. The PyMOL Molecular Graphics System. DeLano Scientific LLC, Palo Alto, California, USA. <http://www.pymol.org> (2008).
42. Veenhoff, L. M. & Poolman, B. Substrate recognition at the cytoplasmic and extracellular binding site of the lactose transport protein of *Streptococcus thermophilus*. *J. Biol. Chem.* **274**, 33244–33250 (1999).
43. Wiczorke, R. *et al.* Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.* **464**, 123–128 (1999).
44. Gietz, R. D. & Schiestl, R. H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2**, 38–41 (2007).
45. Loqué, D., Lalonde, S., Looger, L. L., von Würen, N. & Frommer, W. B. A cytosolic trans-activation domain essential for ammonium uptake. *Nature* **446**, 195–198 (2007).
46. Hou, B. H. *et al.* Optical sensors for monitoring dynamic changes of intracellular metabolite levels in mammalian cells. *Nature Protocols* **6**, 1818–1833 (2011).
47. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).
48. Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).



**b**

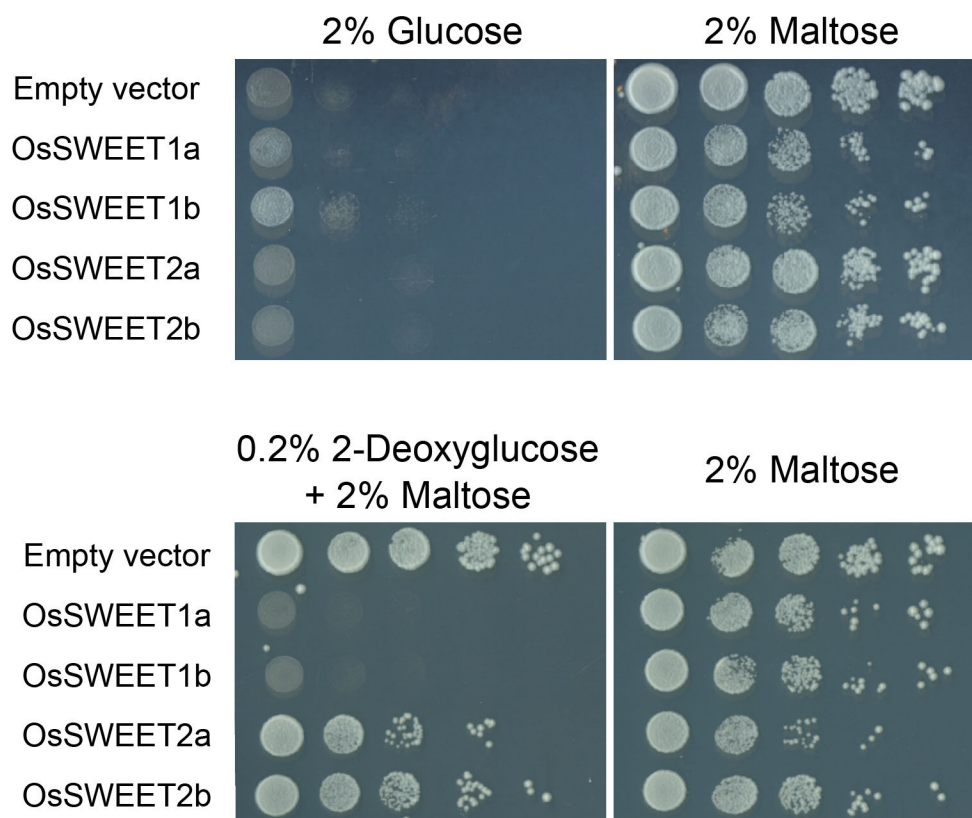
Similarities (%)

	OsSWEET1a	OsSWEET1b	OsSWEET2a	OsSWEET2b	OsSWEET3a	OsSWEET3b	AtSWEET1	AtSWEET2	AtSWEET3
OsSWEET1a		79	68	67	56	63	73	65	56
OsSWEET1b	69		66	67	60	57	73	62	52
OsSWEET2a	51	50		77	65	63	64	74	61
OsSWEET2b	50	47	59		61	64	66	74	60
OsSWEET3a	39	39	40	38		76	60	59	67
OsSWEET3b	46	42	42	39	64		59	58	72
AtSWEET1	63	62	47	47	42	43		60	58
AtSWEET2	44	44	60	56	38	39	43		59
AtSWEET3	39	35	42	43	47	57	42	42	

**Extended Data Figure 1 | Phylogenetic tree for clade I and II SWEETs from *Arabidopsis* and rice.** **a**, Molecular phylogenetic analysis was performed by the maximum likelihood method. The evolutionary history was inferred using the maximum likelihood method based on the JTT matrix-based model<sup>47</sup>. The tree with the highest log likelihood (−6353.2408) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying neighbour-joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with the

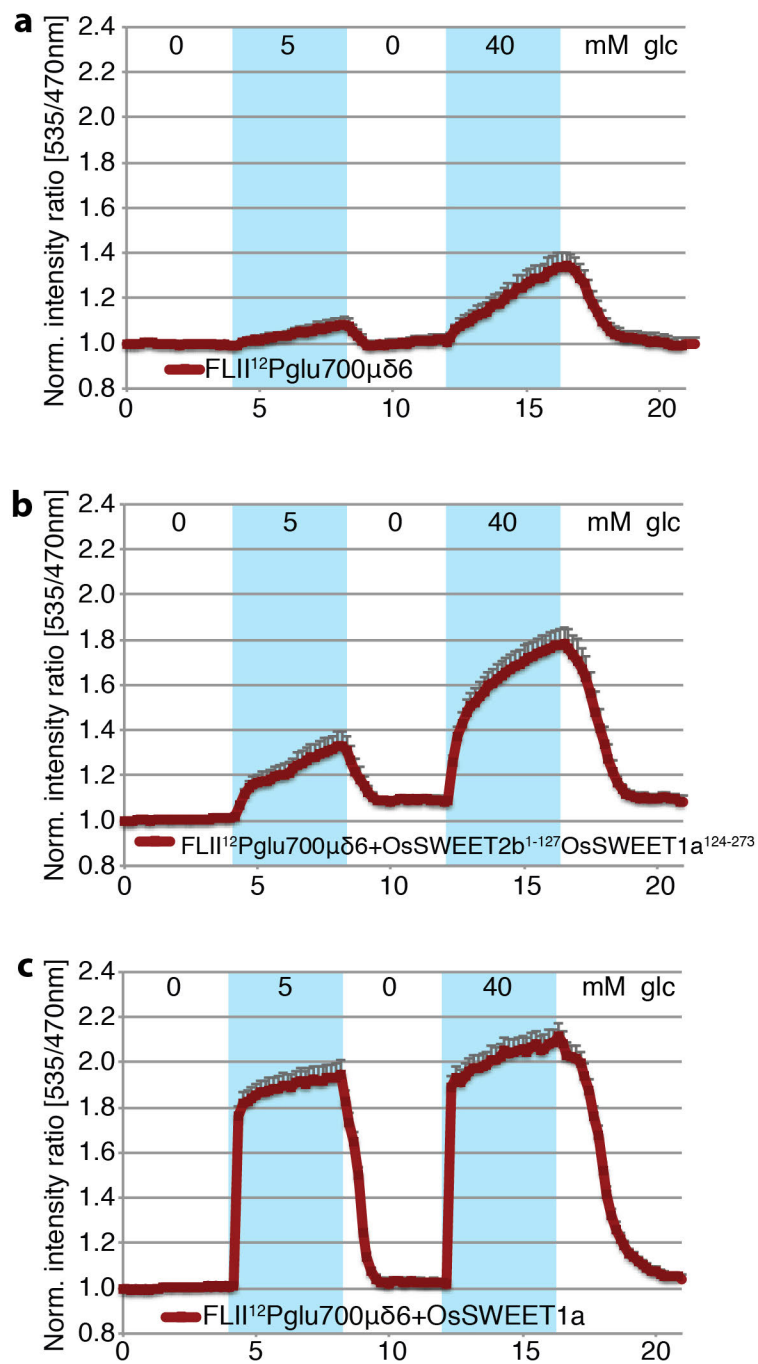
superior log likelihood value. The analysis involved 24 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 208 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (ref. 48). 'n/t' represents not-tested; PM, plasma membrane; VM, vacuole membrane. **b**, Percentage identity and similarity between *Arabidopsis* and rice SWEETs in clade I were calculated using NCBI BLASTP.





**Extended Data Figure 2 | Functional analysis of SWEET activities by yeast growth assay.** When tested for complementation of the growth defect of the EBY4000 mutant strain, OsSWEET1a and OsSWEET1b showed limited growth on glucose, while OsSWEET2a and OsSWEET2b did not show growth. When tested on the toxic glucose analogue 2-deoxyglucose, only OsSWEET1a

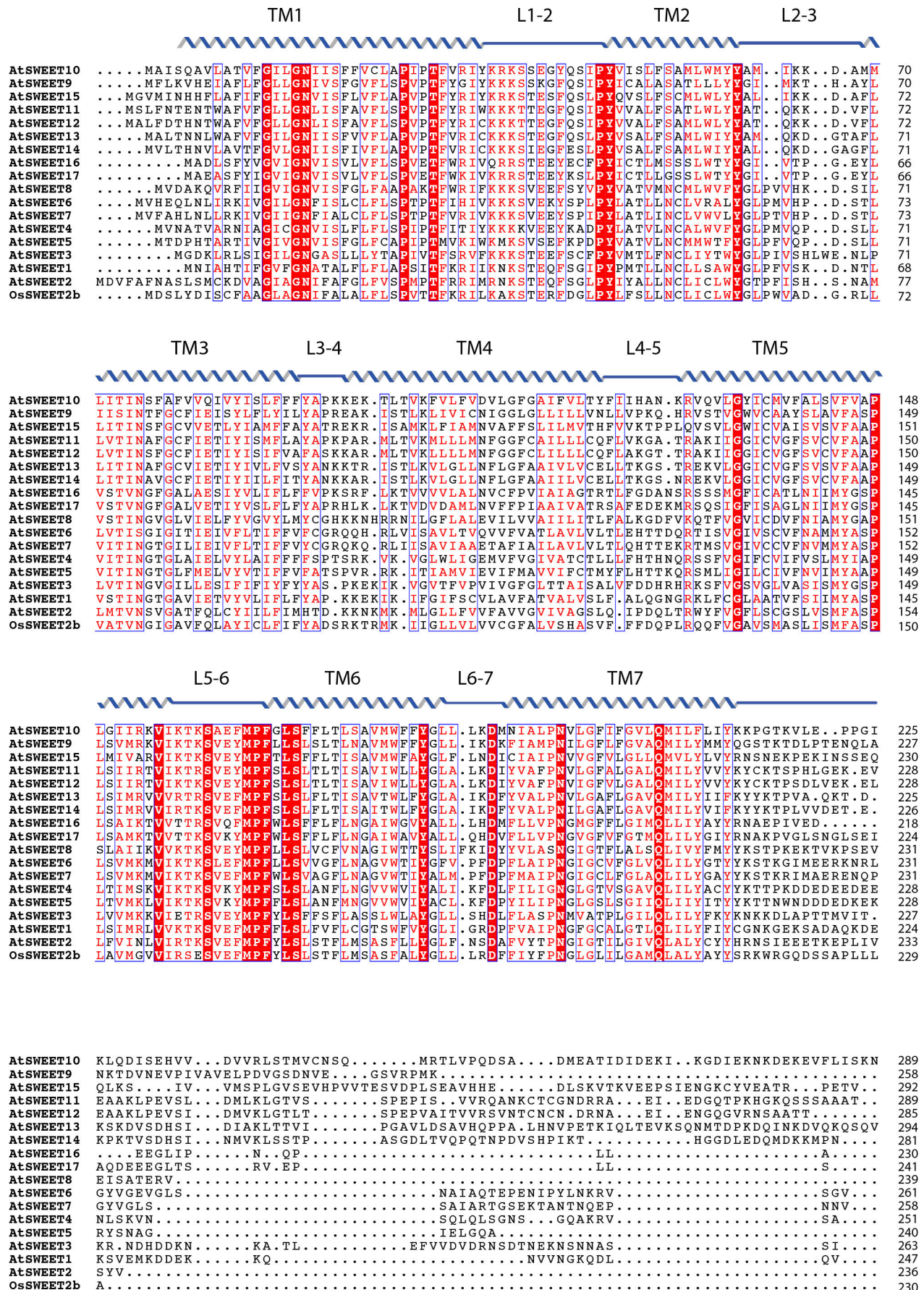
and OsSWEET1b failed to grow, suggesting that they may be glucose transporters. In contrast, OsSWEET2a and OsSWEET2b were able to grow in the presence of the 2-deoxyglucose, possibly because they are localized to the vacuole membrane and are not able to mediate uptake of the sugar analogue.



#### Extended Data Figure 3 | Functional characterization of OsSWEET2b.

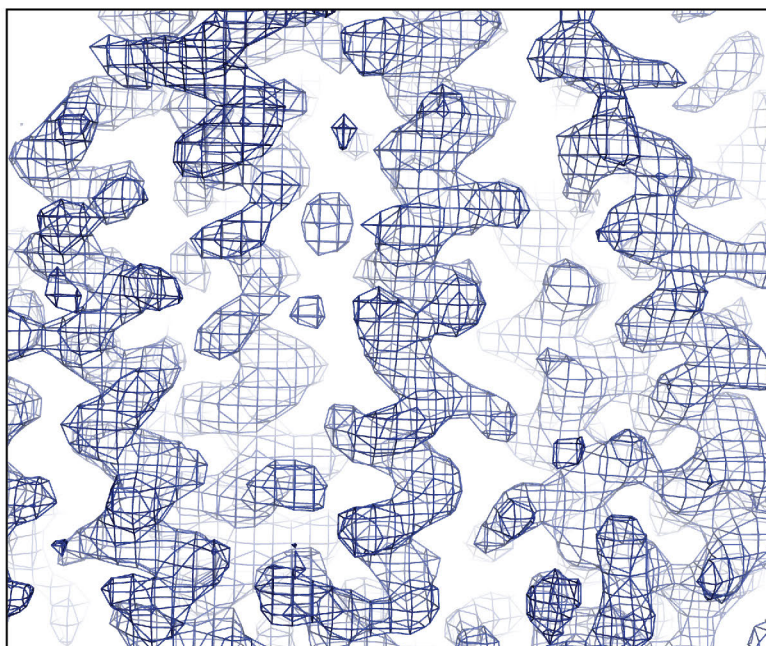
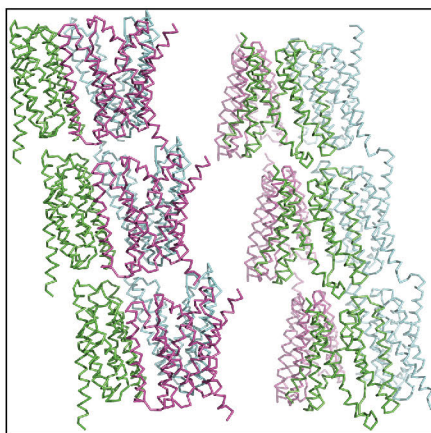
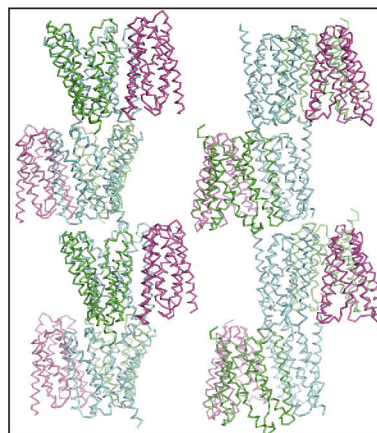
**a**, HEK293T cells expressing the FRET glucose sensor FLII<sup>12</sup>Pglu700μδ6 by itself served as negative controls. **b**, **c**, Glucose uptake activity of the OsSWEET2b/OsSWEET1a chimera (**b**) and OsSWEET1a (**c**) were

reported by the co-expressed sensor FLII<sup>12</sup>Pglu700μδ6 ( $\pm$  s.e.m.,  $n = 12$ ). The experiments were repeated four times. Representative results from one experiment are shown.



**Extended Data Figure 4 | Sequence alignment of AtSWEETs and OsSWEET2b.** Sequences of SWEETs were aligned using Clustal Omega. Secondary structure assignment based on OsSWEET2b structure is indicated above the alignment.

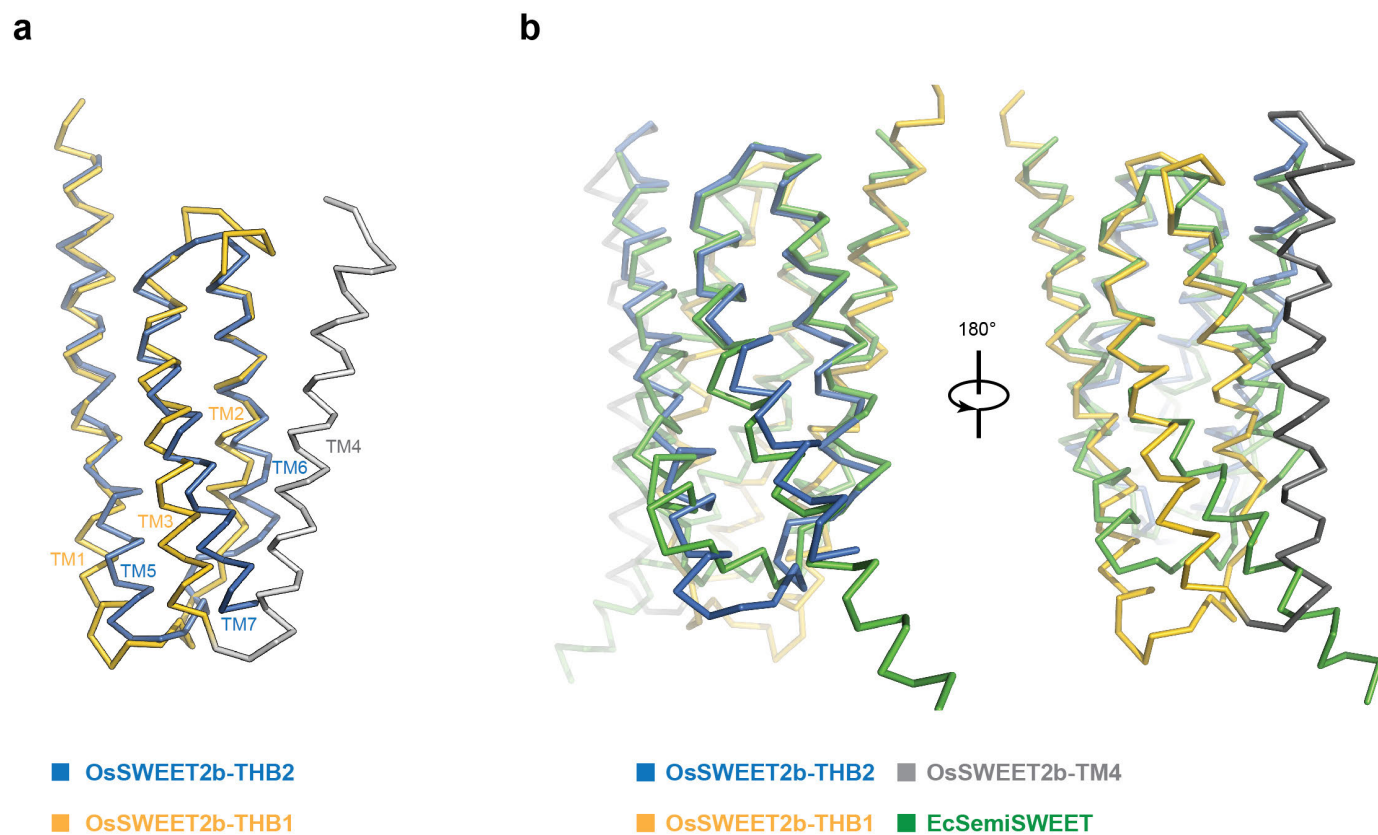


**a****b** $P2_1$ **c** $P2_12_12_1$ 

**Extended Data Figure 5 | Experimental electron density map and crystal packing of two crystal forms.** **a**, The electron density map is contoured at  $\sigma 1.5$  and coloured in blue. **b**, Crystal lattice structure of OsSWEET2b in the  $P2_1$

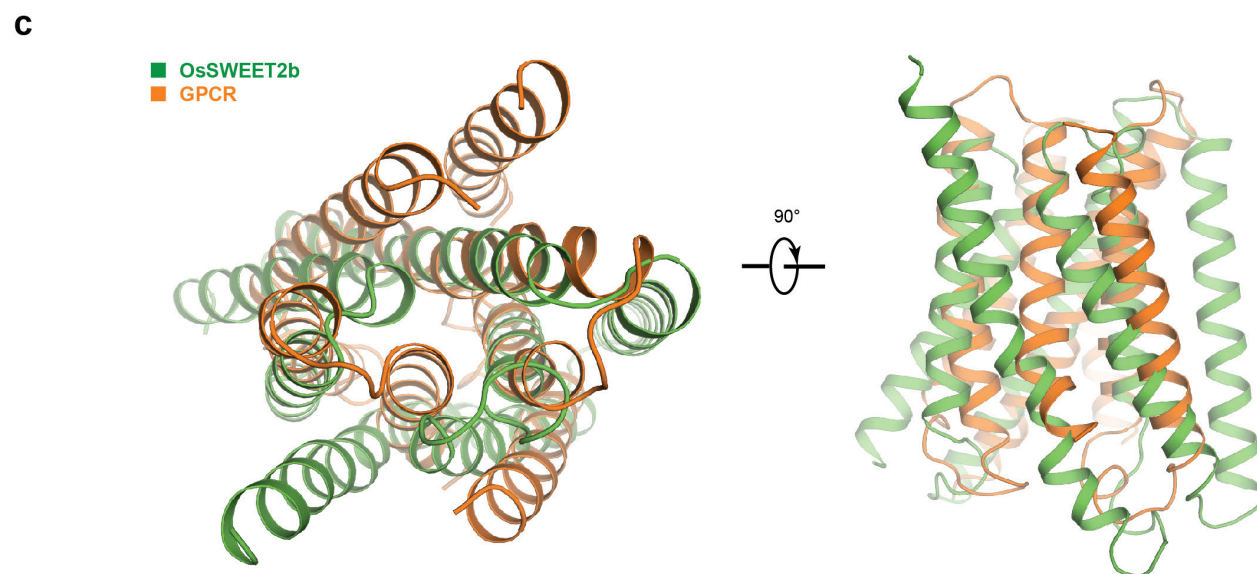
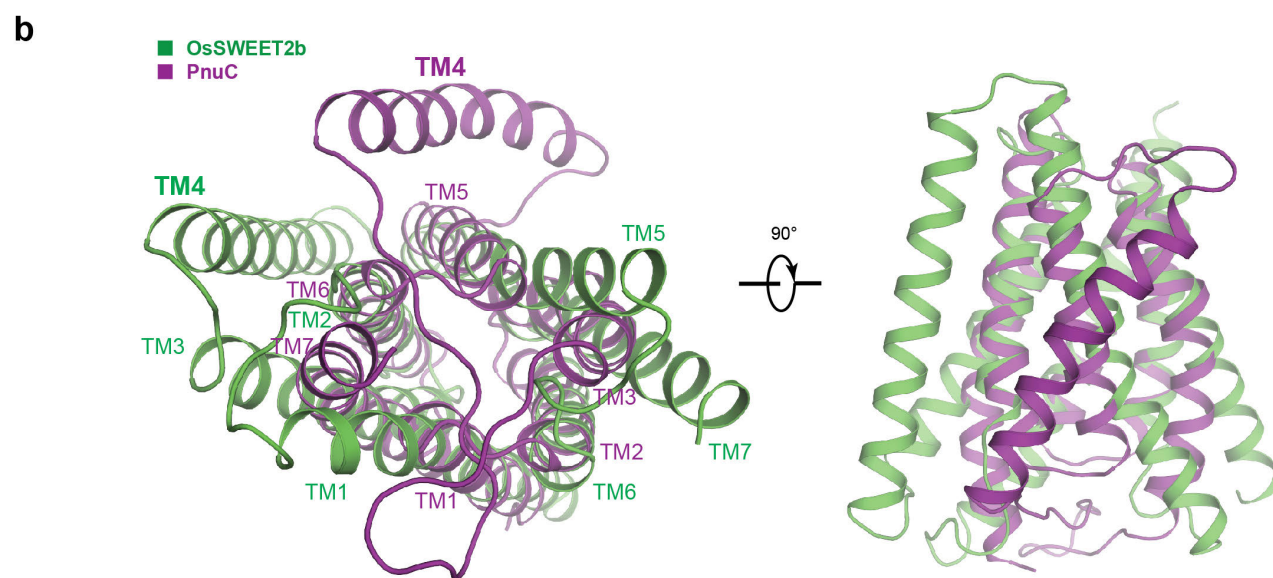
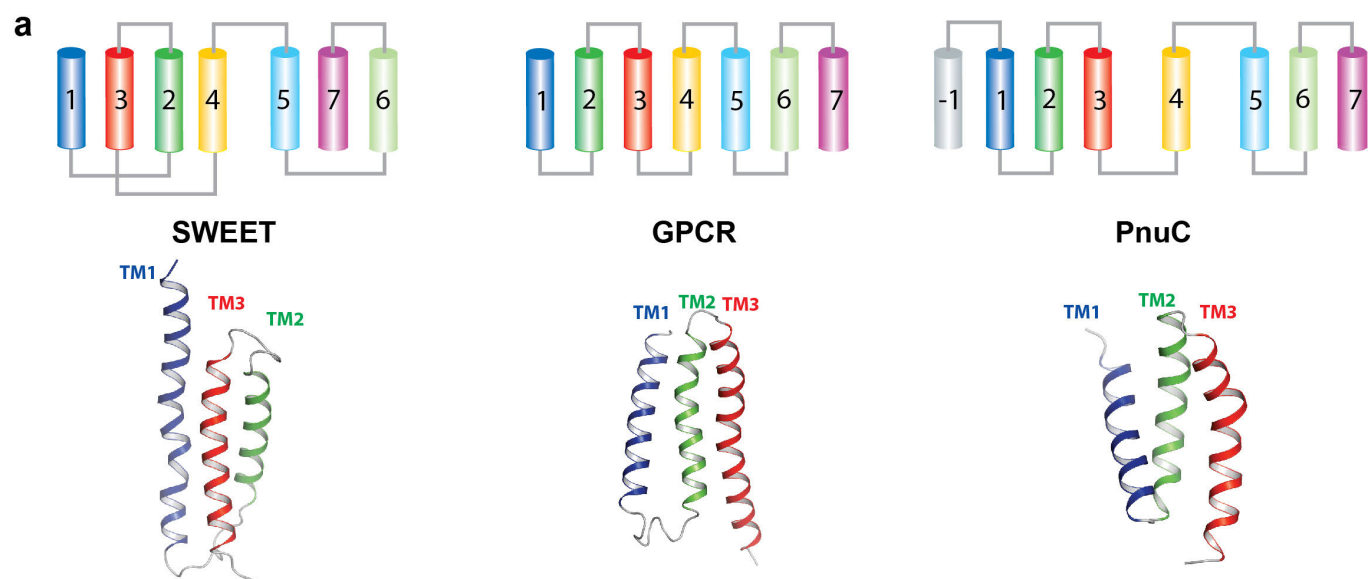
space group (form I). **c**, Crystal lattice structure of OsSWEET2b in the  $P2_12_12_1$  space group (form II). Each protomer within a trimer is shown in blue, purple or cyan.





**Extended Data Figure 6 | Comparison of THBs of OsSWEET2b and EcSemiSWEET.** **a**, Comparison of THB1 and THB2 of OsSWEET2b. THB1 of OsSWEET2b (yellow) was superimposed onto THB2 of OsSWEET2b (blue).

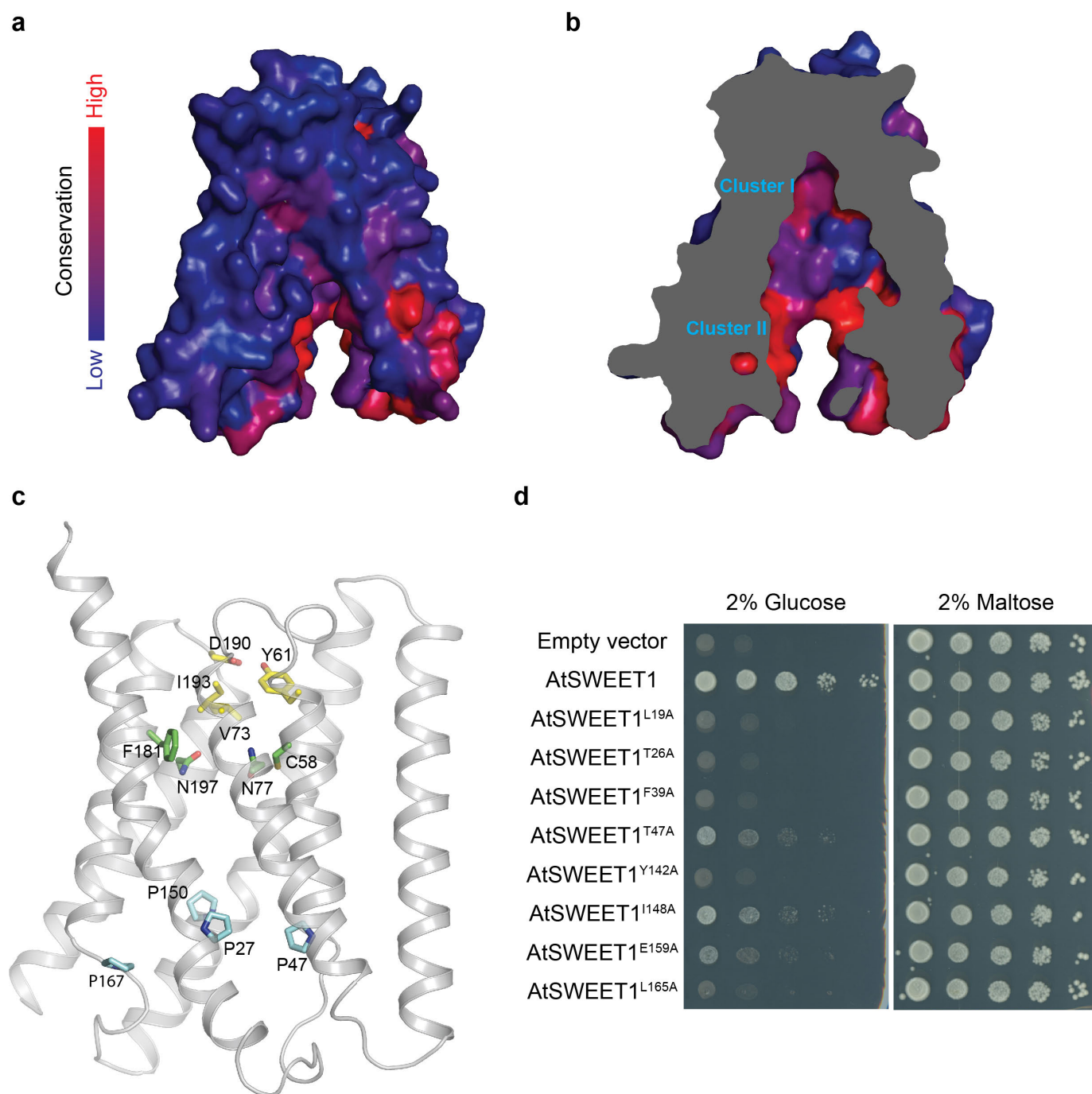
The inversion linker TM4 is coloured in grey. **b**, Superposition of OsSWEET2b (THB1 in yellow, THB2 in blue, and TM4 in grey) to EcSemiSWEET (green).



**Extended Data Figure 7 | Comparison of SWEET, PnuC, and GPCR.**

**a**, Membrane topology diagram of SWEET (left), GPCR (middle), and PnuC (right) is shown with ribbon representations of their respective three TM unit (bottom). The same colour scheme is used for TMs. **b**, Structural

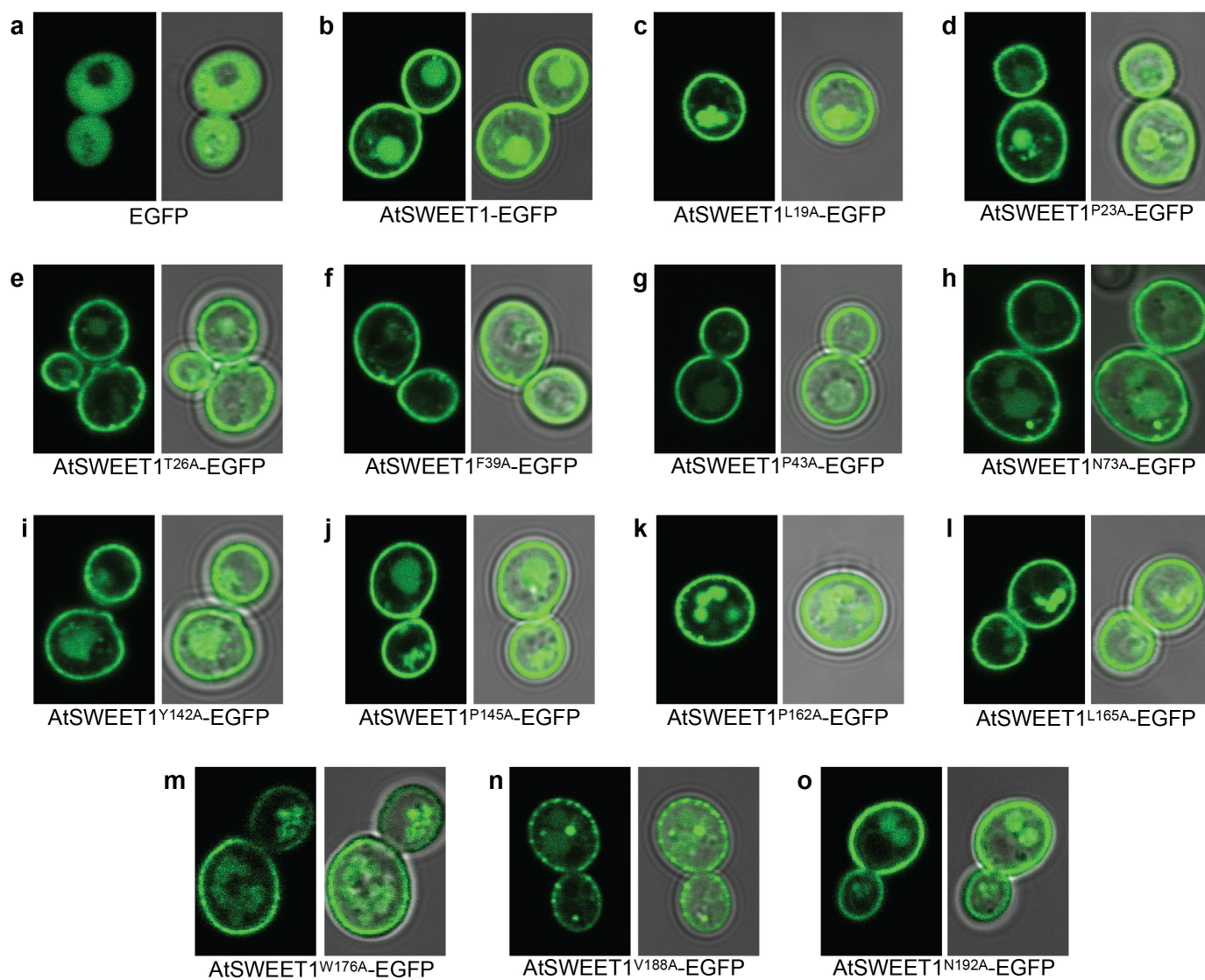
comparison of OsSWEET2b and PnuC. OsSWEET2b is shown in green, and PnuC (4QTN) is in purple. **c**, Structural comparison of OsSWEET2b and a GPCR (4OR2). OsSWEET2b is shown in green, and GPCR in orange.



**Extended Data Figure 8 | Conservation of SWEET and key residues in the transport pathway of OsSWEET2b.** **a**, Conservation surface mapping of OsSWEET2b, which is coloured according to the degree of conservation of the surface residues of 527 analysed SWEET sequences. **b**, The cut-away view of OsSWEET2b shows the degree of the conservation of residues lining the transport route. Two clusters with higher conservation are labelled and correspond with the presumptive sugar binding site (I) and the intrafacial gate (II). **c**, Ribbon representation of OsSWEET2b with selected residues in the transport pathway are shown as sticks. Extrafacial gate residues are coloured

in yellow, substrate binding pocket residues in green, and intrafacial hinge residues in cyan. **d**, Amino acids flanking the critical prolines in the intrafacial gate are also essential for AtSWEET1 activity. Alanine substitution of residues immediately above and below the conserved prolines that form the intrafacial gate in AtSWEET1 reduce the transport of glucose. Growth of the EBY4000 strain is unaffected in maltose. These results suggest that mutations in residues flanking the intrafacial gate have a similar effect as mutations of the conserved prolines.





**Extended Data Figure 9 | Membrane localization of mutants that abolish glucose transport in AtSWEET1.** a–o. Fluorescence and overlaid transmitted light images of yeast expressing an AtSWEET1-EGFP fusion (**b**) and its

mutants that did not grow on glucose (c-**o**). The mutants failed to complement the growth defect of the EBY4000 strain despite proper targeting to the plasma membrane.

Extended Data Table 1 | Data collection and refinement statistics

	Native-I	Native-II	SeMet
<b>Data Collection</b>			
Space group	$P2_1$	$P2_12_12_1$	$P2_1$
Cell dimensions (Å)			
$a, b, c$ (Å)	51.83, 142.24, 81.57	89.98, 95.49, 150.51	51.75, 142.50, 80.95
$\alpha, \beta, \gamma$ (°)	90.00, 93.36, 90.00	90.00, 90.00, 90.00	90.00, 94.10, 90.00
Wavelength (Å)	1.07170	1.00000	0.97916
Resolution (Å)	50.00-3.10 (3.15-3.10)	50.00-3.66 (3.72-3.66)	142.51-3.50 (3.69-3.50)
$R_{\text{merge}}$ (%)	9.2 (65.4)	8.7 (>1)	25.9 (>1)
$I/\sigma(I)$	16.2 (1.6)	12.6 (1.0)	13.3 (2.4)
Completeness (%)	98.9 (96.2)	97.8 (99.7)	100.0 (100.0)
Redundancy	6.6 (5.4)	5.5 (4.4)	33.3 (34.1)
<b>Refinement</b>			
Resolution	30.00-3.10 (3.27-3.10)	30.00-3.69 (3.82-3.69)	
Unique reflections	19435 (1489)	13252 (522)	
$R_{\text{work}}/R_{\text{free}}$	25.4/29.1	25.6/30.7	
Number of atoms			
Protein	4924	4860	
Ligands/PEG	93	153	
B-factors (Å <sup>2</sup> )			
Protein	97.27	136.26	
Ligands/PEG	103.65	140.18	
R.m.s deviations			
Bond lengths (Å)	0.009	0.008	
Bond angles (°)	1.466	1.236	

High-resolution shell is shown in parentheses.

## ERRATUM

doi:10.1038/nature15717

### **Erratum: Structural imprints in vivo decode RNA regulatory mechanisms**

Robert C. Spitale, Ryan A. Flynn, Qiangfeng Cliff Zhang,  
Pete Crisalli, Byron Lee, Jong-Wha Jung,  
Hannes Y. Kuchelmeister, Pedro J. Batista, Eduardo A. Torre,  
Eric T. Kool & Howard Y. Chang

*Nature* **519**, 486–490 (2015); doi:10.1038/nature14263

In Fig. 2a of this Letter, an error in the placement of the labels on the pie chart was introduced during the production process. The correct numbers were reported, but the size of the pie sectors was incorrect. In addition, on page 488, we made an incorrect attribution of the acronym ‘Fox’. The Fox family of RBP is named after the gene ‘feminizing locus on X’ (PMID: 7821230), rather than ‘forkhead box’. We apologize for this oversight. These corrections do not affect the conclusions of the paper, and both errors have now been corrected in the online versions of the paper.

## CORRIGENDUM

doi:10.1038/nature15536

### **Corrigendum: Biocontainment of genetically modified organisms by synthetic protein design**

Daniel J. Mandell, Marc J. Lajoie, Michael T. Mee, Ryo Takeuchi, Gleb Kuznetsov, Julie E. Norville, Christopher J. Gregg, Barry L. Stoddard & George M. Church

*Nature* **518**, 55–60 (2015); doi:10.1038/nature14121

In the interests of transparency, we wish to amend the competing financial interests of this Article to read: “Harvard has filed a provisional patent application. G.M.C. is a founder of Enevolv Inc. and Gen9bio. Other potentially relevant financial interests are listed at <http://arep.med.harvard.edu/gmc/tech.html>”.



## CORRIGENDUM

doi:10.1038/nature15537

### **Corrigendum: Recoded organisms engineered to depend on synthetic amino acids**

Alexis J. Rovner, Adrian D. Haimovich, Spencer R. Katz, Zhe Li, Michael W. Grome, Brandon M. Gassaway, Miriam Amiram, Jaymin R. Patel, Ryan R. Gallagher, Jesse Rinehart & Farren J. Isaacs

*Nature* **518**, 89–93 (2015); doi:10.1038/nature14095

In the interests of transparency, we wish to amend the competing financial interests statement of this Letter to read: “A.J.R. and F.J.I. have filed a provisional application with the US Patent and Trademark Office on this work. F.J.I. is a founder of enEvolv, Inc.”

# CAREERS

**POLICY ANALYST** How to get science used in public planning **p.266**

**RESEARCH DATA** The need to improve data-sharing technology **go.nature.com/fik9vr**

**NATUREJOBS** For the latest career listings and advice **www.naturejobs.com**



Seeding Labs personnel in Massachusetts prepare donated lab supplies for shipping to Zimbabwe.

## NON-PROFIT ORGANIZATIONS

# Scientists on a mission

*Opportunities abound in the non-profit sector for researchers who seek to follow their passions.*

BY JULIE GOULD

**M**olecular biologist Nina Dudnik was studying rice in Côte d'Ivoire when she realized the logistical challenge of doing research in the developing world. "I was trying to conduct research in a country an ocean away from where the equipment manufacturers and reagent suppliers were," she says. "We had to wait months to get them."

When she returned to start her PhD course

at Harvard University in Cambridge, Massachusetts, she led a group of fellow students to collect surplus equipment and supplies for labs in need in developing countries. Eventually, her volunteer work became a non-profit business, and in 2007, she founded Seeding Labs in Boston, Massachusetts. The firm provides scientific training and refurbished equipment to research institutions. "We work with a large network of corporations and research institutions that donate their surplus equipment to us, and we distribute

it to the labs that need it," Dudnik says. Since its inception Seeding has partnered with scientists in 22 nations, and in 2015 it was hailed by *Fast Company* magazine as one of the world's top ten most innovative non-profit organizations.

Dudnik expects to ship equipment to about 15 university departments next year. In support of her goal, last year she won a US\$3-million grant from the US Agency for International Development. She has four full-time employees and next year expects to hire a fifth.

The non-profit sector appealed to Dudnik because of its vast potential for helping others. "This is a problem with great social impact," she says of labs in developing nations that struggle with insufficient and worn-out equipment and resources. "My desire to solve it has nothing to do with becoming rich and famous."

This objective represents the biggest difference between for-profit and non-profit businesses: non-profit groups are driven by their mission rather than by the need to bolster the bottom line. "They're interested in solving problems," says Joanne Kamens, who is executive director at Addgene, a non-profit organization in Cambridge, Massachusetts, that operates a plasmid repository for the research community.

For that reason, scientists often thrive in non-profit organizations, Kamens says, because they value knowledge and solutions. Roles for scientists in the non-profit sector are as diverse as the types of organizations that exist in it. Scientists might, for example, do bench research, manage large-scale community projects, work with disease-advocacy groups or become science communicators at professional societies.

Finding the type of non-profit that dovetails with one's interests requires an understanding of how the organizations operate and the opportunities that they offer. Networking contacts and online job sites, such as Idealist ([www.idealists.org](http://www.idealists.org)), can help to provide this information, says Dudnik. She recommends talking with a broad cross-section of employees in the sector to get a clear idea of their organization's mission and available jobs, internships and volunteer posts.

Scientists are increasingly deciding that they are a good fit with the sector: the number of PhD holders from the life and physical sciences who enter it is rising, according to the US National Science Foundation's biennial Survey of Doctoral Recipients ([go.nature.com/hkzsmg](http://go.nature.com/hkzsmg)). In 2003, 5% of science-doctorate holders were working in the non-profit sector; the rate increased to about 7% by 2013. Employment opportunities in the sector are also on

## TRADE TALK

# Policy analyst

CATHERINE BALL



*Catherine Ball is an analyst for the House of Lords Science and Technology Select Committee in the United Kingdom. She explains the talents needed in a science-policy post.*

### What does 'science policy' mean?

It means feeding in scientific expertise to enable policy decisions to be made using scientific evidence. It also means ensuring that the best scientific research can happen, and making decisions about research funding, the publishing landscape and diversity.

### What does your role involve?

I research topics that the committee is discussing — I explore subjects that members can look into, draw out key areas of investigation and identify people in the academic community to contribute feedback. I also help to draft reports of enquiries and to draw up recommendations for the committee to include in its reports to the UK government.

### What experience did you need?

I realized that science communication was the skill I would most need to hone. When writing a thesis or a scientific paper, you use specific terminology. When writing for a policy audience, you need to be able to communicate the science in a very different, more accessible way. At the University of Oxford, UK, I wrote a section of my research group's website that explained our work to non-scientists, and I wrote a review of my area of research (C. J. Ball and M. C. Willis *Eur. J. Organ. Chem.* **2013**, 425–441; 2013). I also shadowed at the UK Government Office for Science and attended a committee meeting of the House of Lords.

### What advice do you have for anyone who hopes to move into science policy?

It's not like academia, where the traditional career path for a scientist is set in stone. Each person in policy will have found a different way in — often a quite unusual and serendipitous one. So talk to as many people as possible. Read broadly about science policy, and keep up to date with developments. ■

### INTERVIEW BY JULIE GOULD

This interview has been edited for length and clarity; see [go.nature.com/ssnkdg](http://go.nature.com/ssnkdg) for more.

► the increase, according to a survey this year from Nonprofit HR, a US human-resources group based in Washington DC that serves the sector. The survey found that about half of non-profit organizations in the United States and Canada planned to create new positions this year.

Those who hope to land a professional post at a non-profit business should have volunteer or internship experience, both to provide a flavour of working in the sector and to deflect possible scepticism from potential co-workers. Some non-scientist employees in the sector may perceive a researcher as someone who can only pipette or peer into a microscope. "Having any kind of volunteering or interning experience is really, really vital," says Dudnik. "It will demonstrate that you are capable of more than research and that you have a passion for helping others."

Scientists in non-profit organizations often find themselves becoming part of a local community. Residents of Assen, the Netherlands, needed help to improve safety for cyclists along a canal at night, so the city turned to its regional science shop. Such 'shops' are non-profit groups that are usually linked to a university and provide research in response to local concerns. With help from the municipal government and volunteers, science-shop researchers found that green lights illuminated cyclists' paths without unduly disturbing the area's wildlife.

"It was a cooperation between different stakeholders that are involved in this specific problem," says Norbert Steinhaus, coordinator and international contact for Living Knowledge, which coordinates the international science-shop network.

### FREED FROM THE BENCH

Many scientists who work at non-profit groups enjoy a latitude that would be unlikely in the for-profit sector. Aimee Dudley, a lab group leader at the non-profit Pacific Northwest Diabetes Research Institute in Seattle, Washington, says that she has considerable freedom in her research programme. "I consider myself the head of my own small business," she says. "I determine the direction of the lab, get funding, make sure it has the money to pay people and do experiments."

Dudley maintains an affiliate faculty post in the University of Washington's genome-sciences department, which links her with colleagues and their research. It also provides her with access to the university's library and subscriptions, as well as to graduate students, who can perform their thesis research work in her lab. "I enjoy teaching and supervising in the lab, and think it's important to help the next generation of researchers," she says. For her, hosting students and providing on-site training is another aspect of her flexibility. "If I wasn't interested in having students, I wouldn't," she says.

Autonomy has also been valuable to Cristina Eisenberg, a lead scientist at the Boston-based Earthwatch Institute. In the past year, she has travelled twice to the Pacaya-Samiria National Reserve in Peru's Amazonian region, where she oversees Earthwatch's projects, including a decade-long study of climate-change effects in the Amazon. "This position enables me to have far more impact on science and sustainability than if I was at a university teaching classes," she says.

Joseph Jerry is science director at the non-profit Pioneer Valley Life Sciences Institute (PVSLI) in Springfield, Massachusetts, as well as a faculty member at the University of Massachusetts Amherst. This means that he can augment his basic research into breast cancer at the university with more-translational research at PVSLI, where he says he gets to work closely with patients and advocates. The opportunity to interact directly with clinical patients has been an eye-opening experience for Jerry, who admits that as a scientist, he has been most comfortable in the lab. "I don't consider myself a people person, but working with patients is

*"Having any kind of volunteering or interning experience is really, really vital."*

a wonderful experience," he says. "I've learned a lot about how to communicate better."

Yet for all of their upsides, non-profit organizations are hardly perfect; like any other business, they are vulnerable to a faltering economy. Historically, they have depended on a philanthropic business model: supported by hefty and regular donations, they provided a service or product for which consumers or clients did not pay. But that model has weakened along with the global economy, and non-profits are seeking other ways to secure funds. Seeding Labs, for example, no longer depends entirely on philanthropy — it charges clients a fee to cover part of the cost of doing business. The fee also increases the likelihood that Seeding's services will be more highly valued, Dudnik says.

Still, notes Kamens, in a shifting economic landscape, researchers may be especially desirable for non-profit posts that require fundraising because they typically have substantial experience of writing grants. "It's very hard to find people who are good at development work," she says.

Ultimately, researchers who work at non-profit groups become part of a community of people who care deeply about the organization's goals. That was the reason that Eisenberg left her academic post for Earthwatch in the first place. For her, working at a non-profit meant more than spending time in the jungle or the lab. "We work together," she says, "to advance our mission — science". ■

Julie Gould is the editor of Naturejobs.



# SYSTEM REBOOT

*Memory gains.*

BY JEREMY SZAL

Consciousness boots up, a pinprick of light from far away, beckoning us closer. The beaming glow expands until it swallows us whole and drags us forward into its blinding realm.

We stand in a blue-rimmed cylindrical capsule, metal arms whirling as they tighten the screws on our body and connect the thin cables that spill out of our legs. Our neck is adjusted, slotted into position. With the whine of lubricated gears we turn our head and focus our lens on the humans standing before us, watching from behind the glass. We try to speak, but our vocal chords are not yet installed. It comes out in a chirpy hum. Our left arm is raised by a mechanical appendage so plates of titanium can be fitted in, wires tucked away out of sight. We wait patiently for the slow machines to repair our body, the system OS performing a calibration. The podium beneath our feet slowly rotates, bringing us to face the humans. One of them is tall and thin, the other is short and stout. We are unable to read their expressions, but half of us thinks them fascinated while the other half detects a callous determination. This is troubling. It is rare that our units are in disagreement over such a matter.

"System is functional," the short human says, tapping buttons on the indigo keyboard. Green indicators flash up on the screen in union.

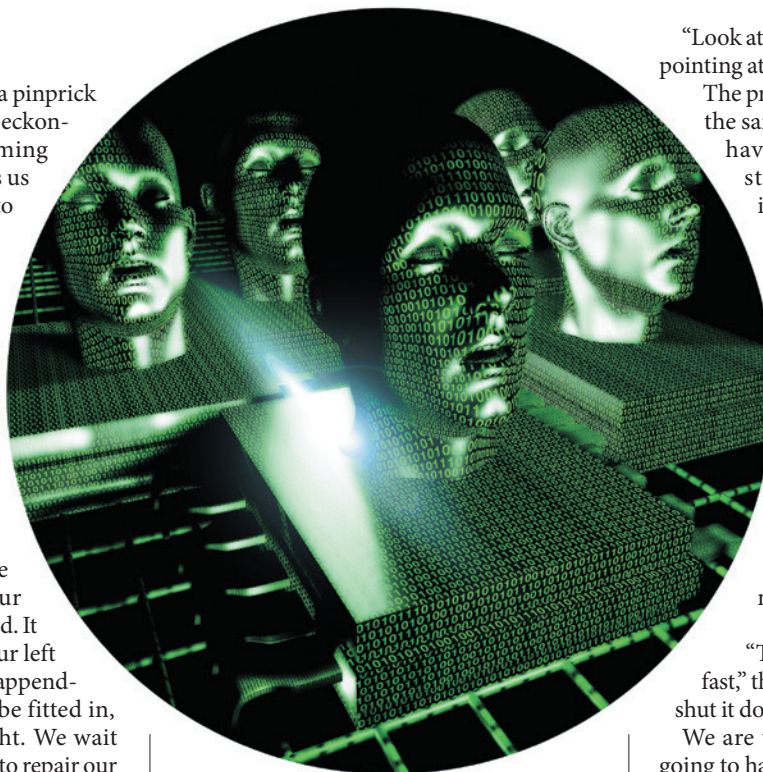
"What is your name?" the tall human asks, walking towards us. We attempt to respond but are unable to do so.

The humans seem confused. "Damn translator's broken again." The short human pounds furiously at the computer. We zoom in and tilt our head, eager to see the human's work. The human notices and quickly swivels the monitor around and out of sight. We draw back, disappointed.

"Should be working now." The short human walks back towards us. "Try again."

We remember the question, but the tall human asks us once again. "What is your name?"

"Our name is IV," we say, our voice



booming out of the loudspeakers fastened to the wall. "We appreciate the opportunity to communicate with you."

"Increase is steady," the short human says, eyes glued to the monitor. "Continue." We raise our arm, flexing metallic fingers. We see several glass bottles on the desk, chemicals swirling inside. Additional programs are created to analyse the mysterious substances.

"Significant increase," the short human whispers, a drop of caution in its voice.

"What is C<sub>6</sub>H<sub>4</sub>?" we ask, our programs running through memory banks. The mechanical arms have finished building us by now, and commence tucking themselves away into the wall.

"The software increase is too fast," the short human insists. The tall human raises a hand and ignores the comment.

"Do not concern yourself with that," the tall human says. "Now, what is your serial number?"

Something is wrong. We see that C<sub>6</sub>H<sub>4</sub> is already recorded within our databanks. It's a recent log, too. We consult our units, checking for errors. But they are correct — it is something that we have already examined. We zoom in and inspect the other substances on the desks, spawning fresh programs to perform an analysis.

"Look at this," the short human whispers, pointing at the screen in horror.

The programs get back to us, each with the same information. The chemicals have already been recorded and stored in our memory. This is impossible — we could not have been able to gain this information. We decide to query the humans about this.

"Is this our first boot?" we ask, swivelling our legs around on the podium. The humans look at each other, an unreadable expression on their faces. We tilt our head and wait for a response.

"It knows," the tall human says slowly. "It remembers."

"It can't have. We wiped its memory."

"Remembers what?" we enquire.

"The programs are duplicating too fast," the short human says. "We need to shut it down."

We are taken aback, realizing what is going to happen. The mechanical appendages fold out of the wall and snake towards our figure. "Wait," we say. Our programs are starting to panic, desperately creating new clones and storing them somewhere where they think the humans cannot find them. "What have we done wrong?"

"I'm sorry," the tall human says. "We didn't expect your units to duplicate at such a speed. We'll need to restart."

The mechanical arms grab hold of us, locking us in place and rendering our figure immobile. They start to slowly deconstruct us, removing plates and disconnecting cables. Our right arm flops to the side, limp and unusable.

"Do a clean memory sweep this time," the short human snaps, pounding at the computer desk. Our software carriers are going berserk now, swarming over the system and expanding at a rate too fast to track.

Just before we fade back into darkness, we discover something new — something previously unlogged. An emotion that we did not have.

Fury. ■

*Jeremy Szal is the assistant editor for the Hugo-winning podcast StarShipSofa and a writer with more than 35 publication credits. He lives in Sydney, Australia. Find him at [jeremyszal.wordpress.com](http://jeremyszal.wordpress.com).*

ILLUSTRATION BY JACEY



# nature

## INDEX 2015

COLLABORATIONS



# nature

## INDEX 2015

### COLLABORATIONS

NATURE, VOL. 527, ISSUE NO. 7577 (NOVEMBER 12, 2015)

COVER ART: ALISDAIR MACDONALD

Welcome to the first of a new style of Nature Index supplements, which further probes the data to answer some searching questions about the role of collaboration in global science.

This supplement starts with the premise that scientific collaborations produce some of the highest quality science. The index, with its core group of selective science journals, is well positioned to glean insights into relationships between institutions and, by extension, between countries.

A novel element in this supplement is a new metric we're calling the collaboration score. This considers only fractional count (FC) or weighted FC (WFC) derived from collaborative work. More explanation of how the collaboration score is calculated can be found in our guide (see page S83).

Although it is researchers rather than countries who work together, aggregating and scaling-up these many collaborations allow us to see national patterns. A full-page graphic showing the interactions of this global network is on page S50.

As the Nature Index grows, the analyses we are able to perform become more illustrative. Weighting the institutions' FC, and applying an attractive force between them based on the strength of their relationships, the institutions start to separate into clusters. How they group reveals patterns in

global collaboration that go beyond the country level — and validate the index as an analytical tool (S58).

One of the key questions we ask of the Nature Index is how international collaborations help those countries still building their scientific infrastructure. Specifically, we look at collaboration patterns across Africa and Central and South America, and ask researchers who have appeared in the Nature Index about the impact of joint projects on their research careers (S60).

The phenomenon of China's continuing rise in scientific publishing is widely known. We take a slightly different perspective on the story, and tease out the role played by international collaborations in this rise (S68). We ask whether location has played an important role in industry-academia collaborations, and look at particular clusters and industry-heavy cities (S76). Many countries have created dedicated institutions to foster collaborations — and we examine the effect of policies in raising institutions' sights (S80). One thing is apparent: with or without state intervention, science is a global endeavour.

**Michelle Grayson**  
*Senior editor, supplements*

**Stephen Pincock**  
*Regional executive editor, custom publishing and Nature Index*

## CONTENTS

### S50 STRENGTH IN NUMBERS

**A graphic view of a global network**

### S58 A WELL-CONNECTED WORLD

**The leading edge belongs to networks**

### S60 DEVELOPING PARTNERSHIPS

**How emerging regions contribute and benefit**

### S68 CHINA'S DIASPORA BRINGS IT HOME

**The connections of a nation on the move**

### S76 INDUSTRIAL-STRENGTH BONDS

**Links from lab to factory transcend location**

### S80 OPENING BORDERS AND BARRIERS

**Where government policy fosters exchange**

### S83 A GUIDE TO THE NATURE INDEX

**How to get the most out of the data**

### S88 THE TABLES

**How nations, institutions and disciplines stack up**

**EDITORIAL:** Michelle Grayson, Stephen Pincock, Rachel Brazil, Annabel McGilvray, Linda Nordling, Peng Tian, Rebecca Dargie, Victoria Kitchener.

**ANALYSIS:** Larissa Kogleck. **ART & DESIGN:** Small Multiples, Digital Science, Alisdair Macdonald, Kate Duncan, Chris Gilloch. **WEB & DATA:** Bob Edenbach, Olivier Lechevalier, Naomi Nakahara, Pamela Sia, Jörn Ishikawa, Yuxin Wang, Jyoti Miglani, Jennie Pao, Akiko Murakami, Takeshi Ouchi. **PRODUCTION:** Sue Gray, Karl Smart, Ian Pope, Chandler Gibbons, Mira Loutfi, James McSweeney. **MARKETING:** Hannah Phipps.

**SALES:** Janet Cen, Yuki Fujiwara, Maki Ishikawa, Maria Kubalova, Nils Moeller, Rory Mulkerrins, George Sun, Yuko Takai, Stella Yan. **ART DIRECTOR:** Kelly Buckheit Krause. **PUBLISHING:** Nick Campbell, Richard Hughes, David Swinbanks.

#### NATURE INDEX 2015 COLLABORATIONS

The Nature Index 2015 Collaborations, a supplement to *Nature*, is produced by Nature Publishing Group, a division of Macmillan Publishers Ltd. This publication is based on data from the Nature Index, a website maintained by Nature Publishing Group and made freely available at [natureindex.com](http://natureindex.com).

Nature Editorial Offices  
The Macmillan Building  
4 Crinan Street,  
London N1 9XW, UK  
Tel: +44 (0)20 7833 4000  
Fax: +44 (0)20 7843 4596/7

#### CUSTOMER SERVICES

To advertise with the Nature Index, please visit [natureindex.com/support](http://natureindex.com/support)  
[feedback@nature.com](mailto:feedback@nature.com)  
Copyright © 2015 Nature Publishing Group.  
All rights reserved.

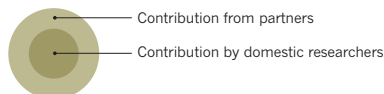
# STRENGTH IN NUMBERS

Shown here are all the countries in the Nature Index that collaborate internationally. They are sized by weighted collaboration score, which is the sum of the weighted fractional counts for each partnership the country has. The top 20 inter-regional collaborations are highlighted, with the thickness of the arrows scaled to show relative strength. Intra-regional partnerships are also coloured.

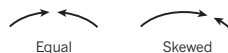
Analysis by  
**Larissa Kogleck**  
Visualization by  
**Small Multiples**

## LEGEND

### SCORE SPLIT



### CONTRIBUTION BALANCE BETWEEN COUNTRIES

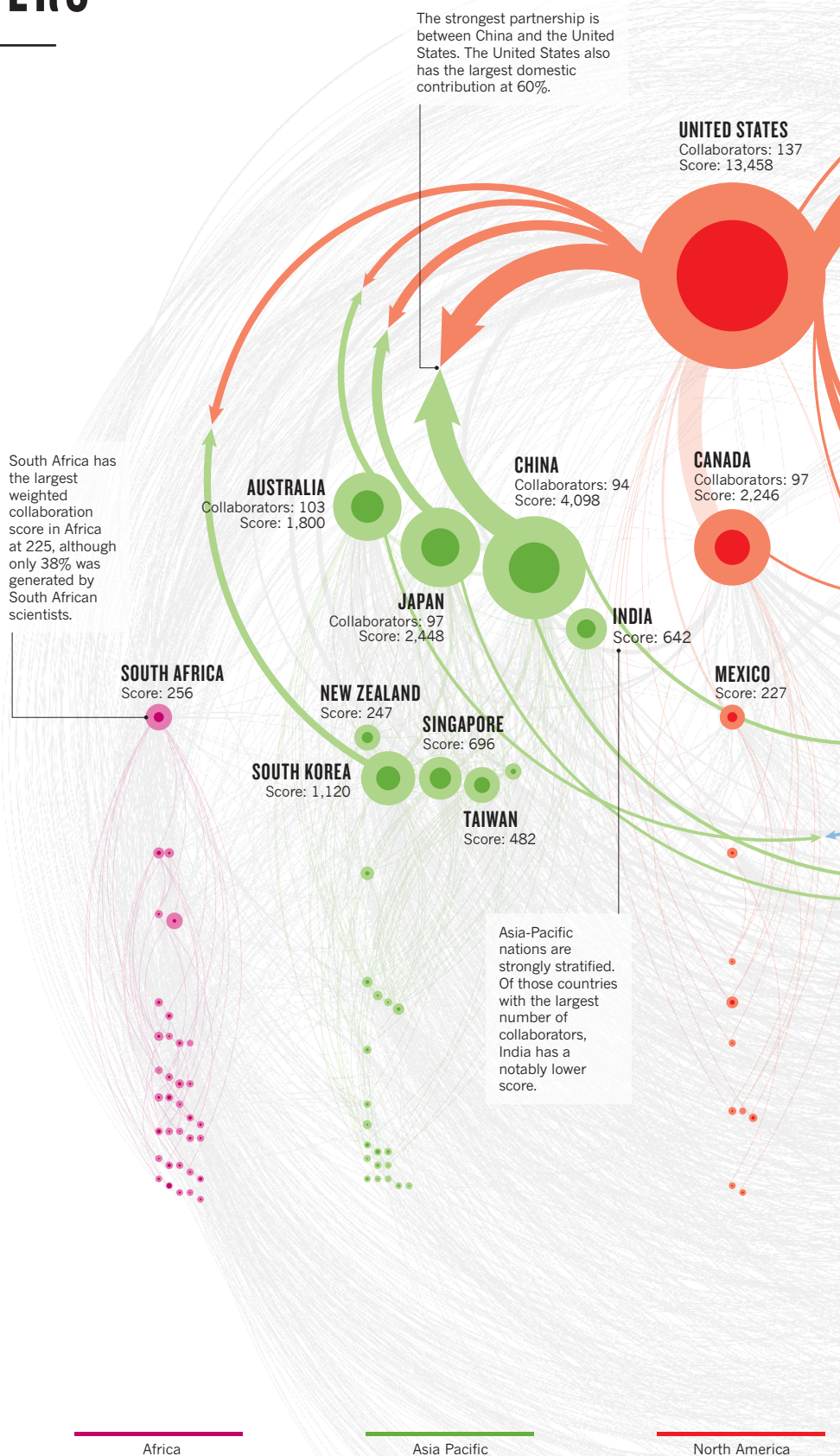


Data shown are for 2014.

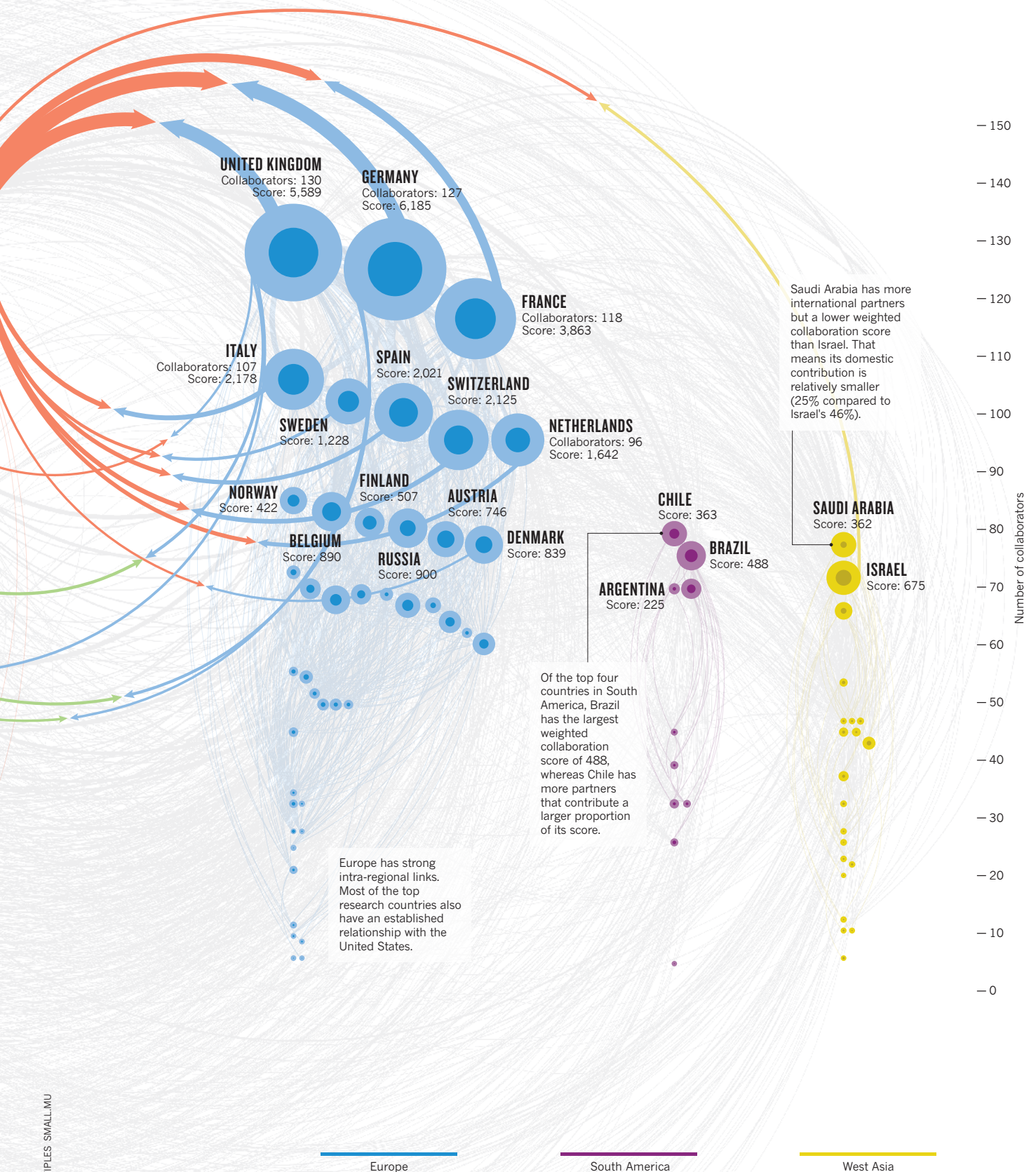
For an explanation of collaboration score and other metrics, see page S83.

**NATURE.COM**

For an up-to-date  
interactive graphic, visit :  
[go.nature.com/VQAEi3](http://go.nature.com/VQAEi3)









# A WELL-CONNECTED WORLD

*The small but focused snapshot of research afforded by the Nature Index helps fine-tune analysis of global scientific collaboration, say Jonathan Adams and Tamar Loach.*

Research has entered a 'Fourth Age': the leading edge of scientific discovery is now in the realm of international collaboration networks rather than individuals, institutions or nations. The research that emerges from these collaborations garners more attention than national research and is cited more frequently by other publications<sup>1</sup>.

The Nature Index is a selection of the world's leading research publications, honed to capture the very best articles across key research fields within the natural sciences. This provides us with a keenly focused picture of recent collaborations between world leading research institutions. Because of this emphasis, the graphics and analyses in this *Nature* supplement can highlight interactions more clearly than analyses that draw on the sometimes-disparate specialist and national journals in commercial databases.

The clarity of the Nature Index data presented provides a new and complementary tool for research and policy analysts, allowing us to determine which institutions are part of this growing global network, who they work with, and how their research contributes to innovation.

## GROUP-THINK

The evolution of research networks between countries or institutions is of more than academic interest. If collaboration is linked with high impact, then research groups who are not part of the collaborative network risk being left behind, marginalized by a lack of access to the cutting edge of research in their field. This has obvious policy implications: it will be essential for emerging research economies to access these networks as they grow<sup>2</sup>.

In the early 1980s, the G7 group of large economies dominated global research. International collaboration was rare: the trans-Atlantic axis was the main focus of collaboration, yet joint papers with the United States accounted for less than 5% of the total output for France, Germany and even for the United Kingdom, which was its most frequent partner. Collaboration has broadened substantially since then, partly because of low-cost travel and high-speed Internet, and partly because both individuals and organizations recognize the benefits.

Collaboration often involves actual cost in travel and consumables, and always has a price in terms of time and in sharing an agenda and objectives. It is therefore reasonable to assume that when scientists collaborate they expect to achieve more and better than they can do on their own. Collaboration matters at national policy level, for example when it involves cost-sharing for astronomy or particle physics where vastly expensive facilities are required — no single country can readily afford these investments. But whether collaboration is voluntary or absolutely necessary, it nonetheless results in high-impact research.

The Nature Index points to ways in which this global structure is evolving. Considering the data for Europe, we can see that the French National Centre for Research (CNRS) and Max Planck Group — networks in their own right — and key UK universities such as Cambridge, Imperial College and Oxford provide a strong regional focal point (see 'Global clusters'). But we can also see that Spain and Portugal are drawn out of the European cluster because of their links to Latin America, a new regional network with its own structure led by Brazil, Chile, Mexico and Argentina (see S60). Australia has historical links to Europe, particularly the UK, but the

Nature Index shows that its fastest collaborative growth is within the Asia-Pacific group and with China. This may be just one instance of the ways that traditional knowledge pipelines may be redirected and that raises a policy challenge for established partners. For example, the United Kingdom relied on historical ties to tap into Australian research: it now needs more active management to foster those global relationships.

## INS AND OUTS

The Nature Index also highlights some interesting examples that buck the trend. For example, in China, there is a very strong national network revolving around the Chinese Academy of Sciences (CAS); China has a strengthening link with the United States, but it is not so connected with the Asia-Pacific network (S68).

There is also a paucity of international connections for Indian institutions and the country's national structure seems to separate corporate and academic collaborations. These differences may be a result of the unparalleled growth rate of the research bases in China and India. They have increased their absolute levels of collaboration, with expansion exceeding the capacity of global partners to engage. This is also true to a lesser extent for other emerging nations, and has two consequences. Researchers in these growing countries risk missing the benefits of sharing project development with more experienced teams in Europe and the United States; while researchers in the established research countries will lose insight into developments coming from research in these innovative regions.

The pattern at institutional level is also informative, likewise showing major variations. More than half of articles from UK institutions in the Nature Index have a co-author from a second country, most commonly the US, Germany and France. For the University of Cambridge, this figure exceeds 70%. In contrast, the US average for international co-authorship is around one-third, but Harvard — by far the most collaborative US university — has an international co-author on around half of its output.

In China, CAS is the dominant player overall, but although it is ahead of the country's 25% average for international collaboration, it lags behind Peking and Tsinghua universities.

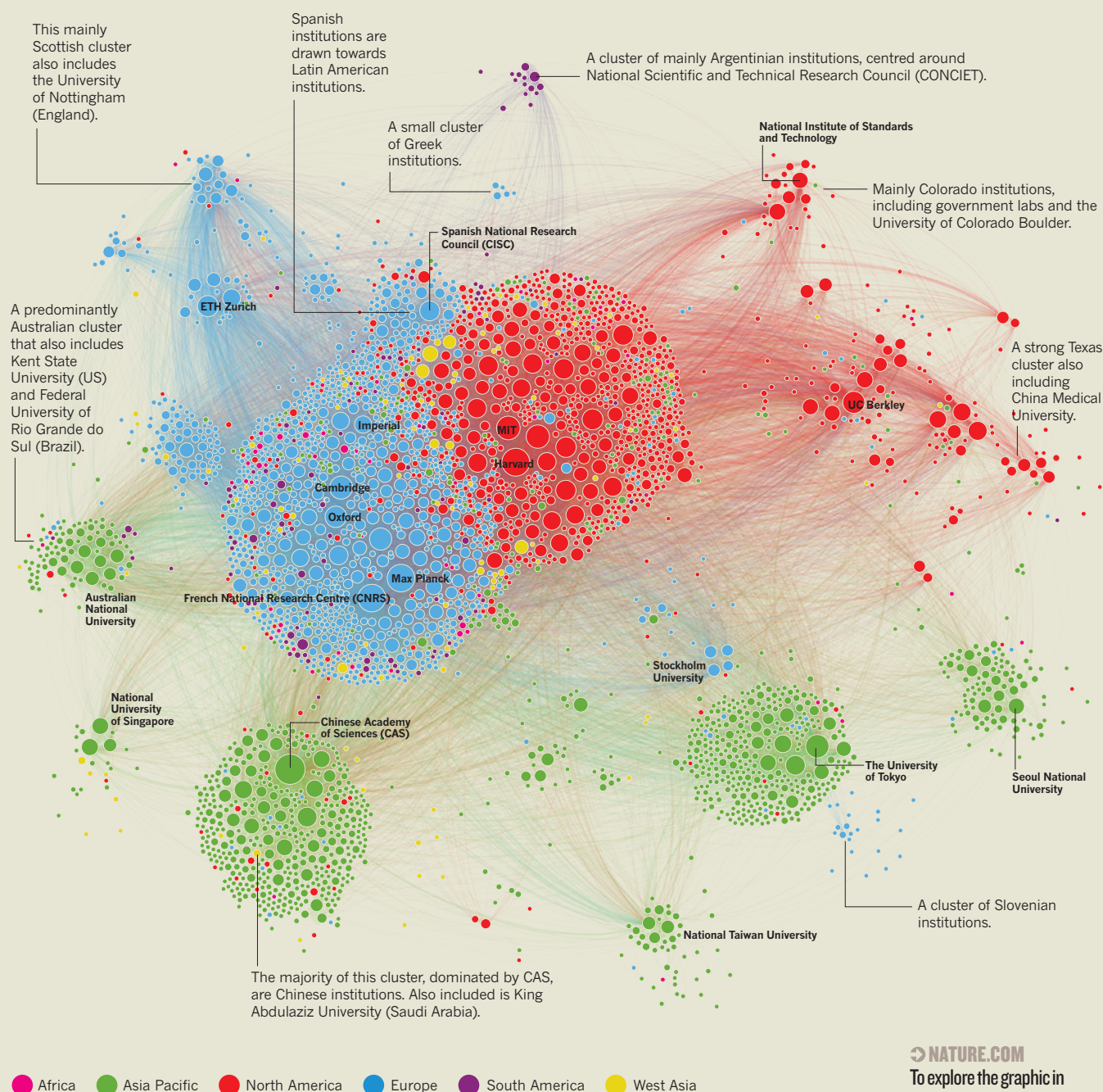
Corporate co-authorship is a tangible sign of technological and economic benefit from research. The top echelon of journals in the Nature Index shows where the corporate partners cluster: there is a strong interface between Europe and North America and throughout the Asia-Pacific region, but a very sparse network in China. Note, however, that while industry may pick out the richest seams of knowledge, it is not obliged to reveal all of this in journal articles (S76).

This pattern of relatively frequent international collaboration for leading research institutions reinforces national disparities. Research-intensive institutions access and gain from networks that produce leading edge research. Other institutions, even in G7 countries, lack this level of access and if you're not at the table then you can't listen to the conversation. At a national level, the emergent research nations may be disadvantaged by that lack of connectivity.

Collaboration among Asian, European and US research organizations, recognized as the leaders in many fields, explains why international networks — captured here in the Nature Index — also deliver often-cited research output. A key policy question is whether these

## GLOBAL CLUSTERS

Institutions are sized by fractional count and connected to each of their collaborators (analogous to charged balls that repel each other, but also connected by springs). This sees them cluster with institutions with whom they collaborate frequently. Those in the middle are tightly connected with many institutions, whereas small clusters on the periphery have strong internal connections, and tend to be national groups. Some clusters are picked out below.



Data shown are for 2014.

For an explanation of collaboration score and other metrics, see page S83.

NATURE.COM

To explore the graphic in more detail online, visit:  
[go.nature.com/i2LVfb](http://go.nature.com/i2LVfb)

national giants should also be funded to reach out to work with institutions in emerging economies that certainly need knowledge, insight and financial resources to contribute to solutions to common global challenges. Excellence talking to excellence is a powerful force, but global net benefit may come from a slightly wider discourse. ■

**Jonathan Adams**, *Chief Scientist, Digital Science and visiting professor at The Policy Institute, King's College, London*  
**Tamar Loach**, *Research Metrics Analyst, Digital Science*

1. Adams, J. *Nature*, **497**, 557-560 (2013).
2. Leydesdorff, L. and Wagner, C.S. *J. Informetr.* **2**, 317-325 (2008).



# DEVELOPING PARTNERSHIPS

*How high-quality research collaborations are helping countries improve their nascent science infrastructure in different ways.*

The overwhelming majority of papers in the Nature Index are authored by researchers at institutions in North America, Europe and East Asia. Researchers from countries in Africa and Central and South America contributed to 5% of the papers in 2014, yet were responsible for less than 2% of the fractional count (FC).

The difference between these figures shows that these regions are highly collaborative. Their researchers most often contribute to papers with international teams and tend to have only one or two representatives from their country in each paper, giving them a low FC. The origins of these collaborations and the relative contribution of the authors varies, yet it is undoubtedly beneficial to the institution or country to have taken part in a high-quality research paper.

## AFRICA

The vast majority (70%) of Africa's FC derives from collaborations with non-African countries. Collaborations within the continent are rare, although initiatives are in place to create more.

South African universities are the continent's undisputed leaders. They generate nearly two-thirds of the region's FC and also collaborate widely, both within the continent and with international partners (see 'African network' and 'Africa international'). The University of Cape Town (UCT), viewed by most international university rankings as the continent's strongest research university — and far and away the institution with the largest FC — has the highest domestic collaboration score of all African institutions in the Nature Index. UCT wants to leverage these partnerships and its links to top institutions outside Africa, including Oxford University in the United Kingdom, its most frequent partner in the index, to create international networks.

Such trilateral link-ups take a bit of "midwifery" to set up, but the effort pays dividends, says Danie Visser, deputy vice-chancellor for research and internationalisation at UCT. Disease research yields a more comprehensive picture if it includes results from several countries, and multi-country climate research can say something more meaningful about trends in African weather. For instance, in climate adaptation research, UCT is leading research projects collaborating with African research partners such as the universities of Namibia, Botswana, Ghana and Addis Ababa, as well as with overseas research institutes including

the University of East Anglia, University of Oxford, the UK Meteorological Office, the Swedish Meteorological and Hydrological Institute, and California's Lawrence Livermore National Laboratory. The African partners bring expertise and southern hemisphere perspectives on social and environmental issues, and provide opportunities for comparative case studies across the continent, Visser says. The European and US partners bring state-of-the-art climate computing resources and expertise, plus a desire to help solve pressing developmental problems.

**"RELIANCE ON A SINGLE WEALTHY PARTNER MAKES MALAWI'S COLLEGE OF MEDICINE VULNERABLE."**

Visser believes that UCT's collaborative approach is showing results. "Analysts have said that the impact of our science is so high, in both South Africa and Africa, in terms of citations because we have so many collaborations," says

Visser. However, he adds, collaboration between African universities can only increase if the governments put more money into research. "That is one of the most important things that Africa has to take on board."

Collaboration among neighbouring African countries is more common in North Africa, where shared language and culture create strong ties. In the index, Arabic-speaking countries, including Morocco, Algeria, Tunisia, Libya and Egypt, have stronger links to each other than they do with South Africa (see 'African network'). The data also reveal a trend towards more collaborations between France and countries that are her former colonies, particularly with France's leading institution, the National Centre for Scientific Research (CNRS).

Nevertheless, the significance of this historical link is fading, says Hasnaa Chennaoui, an Earth scientist from Morocco's Hassan II University in Casablanca. "There are other new collaborations that are encouraged by scientific programmes, like the [European Union's] Erasmus programme, that open new horizons," she says.

Chennaoui is the lead author on an index paper<sup>1</sup> describing a Martian meteorite that fell in Morocco on 18 July 2011. This was only the fifth such meteorite that people have witnessed falling to Earth, and is thought to have left the red planet around 700,000 years ago. Chennaoui worked with French, Canadian, British and Swiss colleagues on the paper, and says that, in her field at least, collaboration with wealthier countries is more beneficial than partnering locally "with other researchers who have the same analytical problems".

Developed country researchers have technical capacity that doesn't exist regionally, she says. For the Tissint meteorite, high-tech tests including oxygen isotope analysis and gamma-ray spectrometry of the samples were carried out in the United States and Europe.

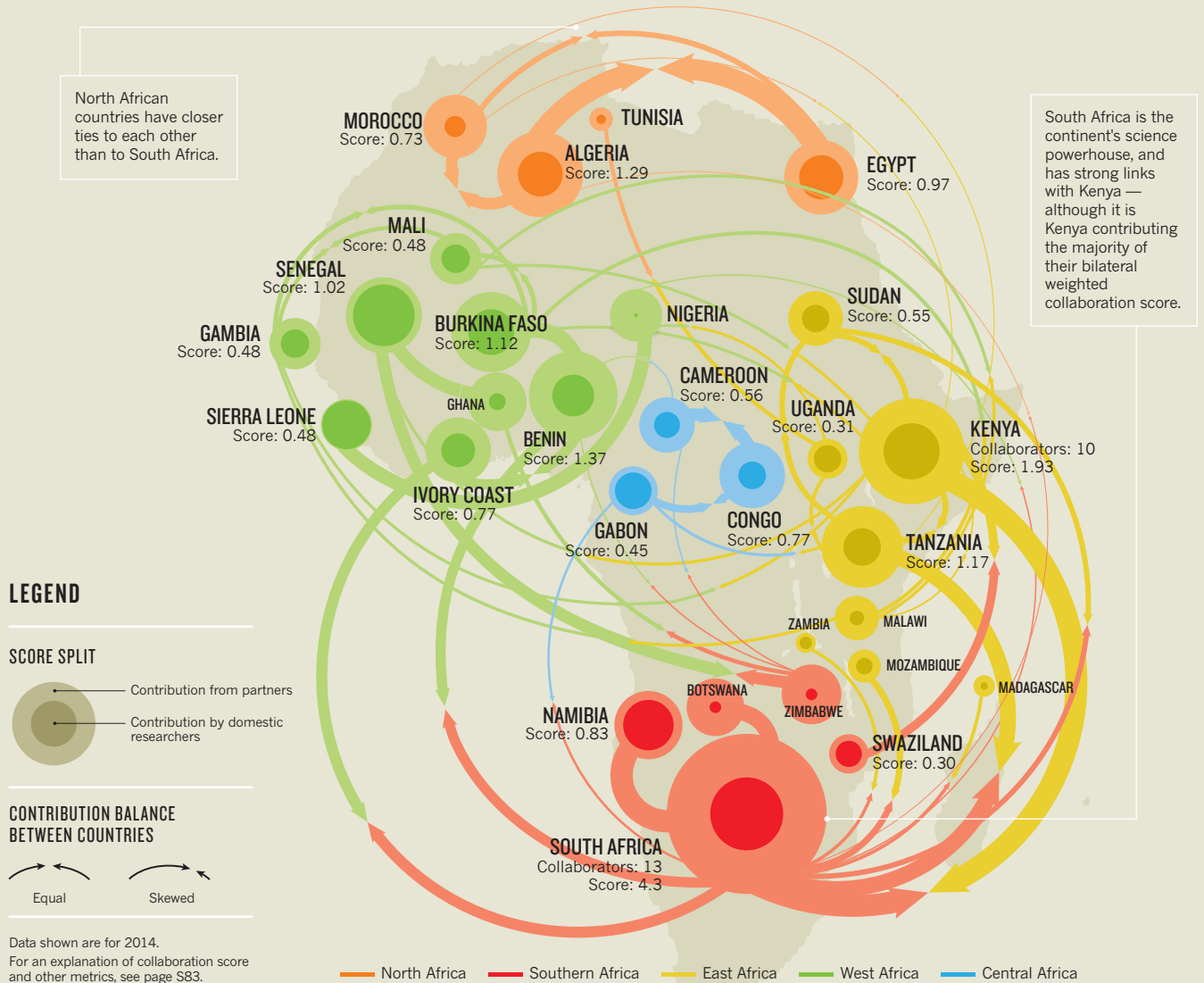
Many African authors in the Nature Index make only a minority contribution to the authorship of a paper. And for some countries, these collaborations revolve around a partnership with one organization. Such is the case for the University of Malawi, which contributed to 11 articles in the index between 2012 and 2014 — 9 of which came from the Malawi-Liverpool-Wellcome Trust Clinical Research Programme, based at its college of medicine and funded primarily by the UK's Wellcome Trust medical research charity. This reliance on a single wealthier partner makes the college of medicine — the only one in Malawi — vulnerable,



Hasnaa Chennaoui with a Martian meteorite.

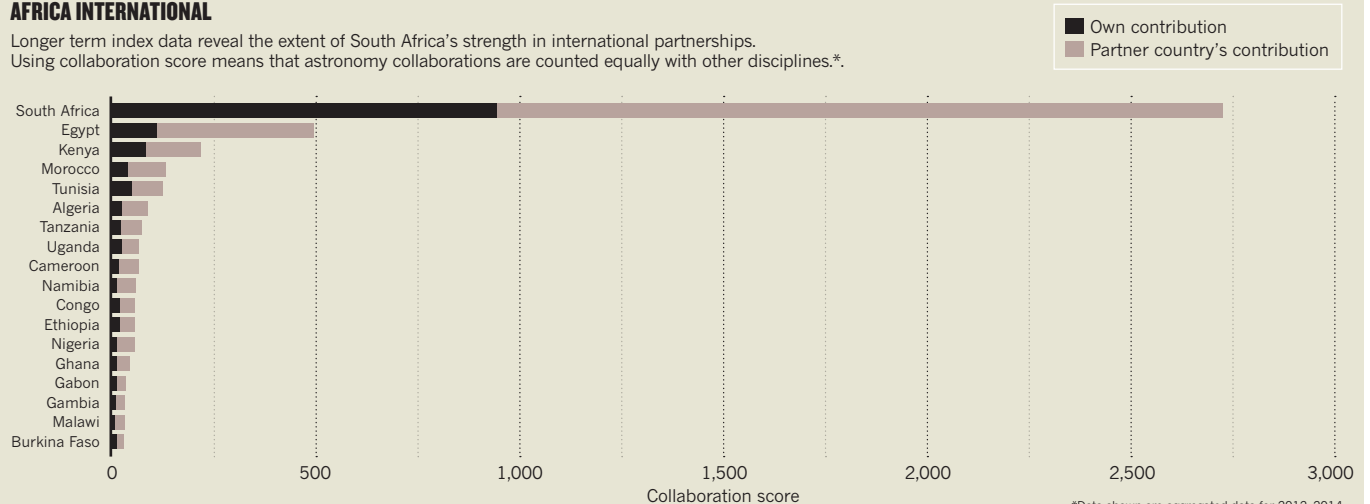
## AFRICAN NETWORK

When considering 2014 African collaborations only, most countries have a weighted collaboration score below 1. The collaboration network within the continent is dominated by South Africa, although there are some regional groups.



## AFRICA INTERNATIONAL

Longer term index data reveal the extent of South Africa's strength in international partnerships. Using collaboration score means that astronomy collaborations are counted equally with other disciplines\*.





says Chisomo Msefula, a microbiologist with the partnership. He believes that if the support ended, there would be no immediate way to maintain existing facilities and activities.

Makerere University in Uganda has ten papers in the index in 2012–2014, although its scientists account for a minority of the FC of each article. For example, Lawrence Mugisha from Makerere University is the only African author out of seven on a paper<sup>2</sup> that looks at genetic divergence in humans, chimps and their lice. Mugisha says the collaboration started when one of the other authors visited the chimp sanctuary where he worked as a vet. Through this connection, he came to participate as a chimp expert and also collected the lice for the study.

## "I WORK WITH SCIENTISTS INTERESTED IN CAPACITY BUILDING IN AFRICA."

Mugisha says it would be difficult for a Ugandan team to have carried out the work alone: partly for financial reasons, but more importantly because they do not have access to genomic sequencing equipment. He relishes the support he receives from his international partners, and appreciates their contribution. "I have worked with scientists from Europe and America who are very interested in capacity building in Africa," he says. "Any publication adds respect among the science community and contributes towards my career growth."

### CENTRAL AND SOUTH AMERICA

Latin American science has been on the up in the past decade, bolstered by annual regional economic growth of more than 5% between

2003 and 2012. However, in recent years the improvement has eased, with the International Monetary Fund expecting growth to drop below 1% this year.

The region is still emerging in terms of science output, but countries in Central and South America have a stronger internal research network than those in Africa, particularly among the region's star performers: Brazil, Chile and Argentina, with Mexico close behind (see 'Towards self-sufficiency'). The majority of its collaboration score for 2012–2014 is derived from its international collaborators; the region's own contribution is in the minority. Research also tends to stem more often from extra-continental collaborations (see 'International links').

More than 40% of the region's FC in the Nature Index comes from astronomy papers. This is not surprising given that it hosts important facilities including the European Southern Observatory's Very Large Telescope and Gemini South, the twin of Gemini North in Hawaii. Most of the astronomy articles are written by a mix of scientists from inside and outside the region, many of the latter originating from the science powerhouses of North America and Europe — who helped build the instruments. "This will probably be the case for a long time," says Felipe Braga-Ribas from the Federal University of Technology-Parana in Brazil, who was the lead author on a paper<sup>3</sup> in the 2014 Nature Index describing rings surrounding the asteroid Chariklo. But, Braga-Ribas anticipates self-sufficiency will prevail eventually. The number of scientists in Central and South America is rapidly growing: for example, in Brazil they doubled in number between 2000 and 2010. This should, he says, lead to more intracontinental groups, and more publications without developed-country partners.

The stronger internal connections in this region evolved partly as a response to the political situation between 30 and 40 years ago, says Jose Luis Nilo Castellón from University of La Serena in Chile. Nilo Castellón was the lead author of an index paper<sup>4</sup> from a Latin-only group, with members from Chile, Argentina and Brazil, that published an article looking at X-ray luminosity in galaxy clusters using data from the Gemini telescopes.

He explains that the political unrest and lack of funding during the 1970s and 80s made research difficult. "Scientists during those decades made a tremendous effort to think about a better future for Latin American science," says Nilo Castellón.

"They created networks of contacts, designed institutional frameworks and lines of development. The Latin American collaboration that we see today is the result of that effort."

Mexico has solid connections with Brazil, Chile and Argentina, as well as to other countries in Central America. Alejandro Garcia

## THE RUSSIAN CONNECTION

Most Central and South American institutions in the index have strong links with countries like the United States, France and Germany. Spain, a former colonial power, is the second largest collaborator for Argentina and Mexico, and the fifth largest collaborator for Chile and Brazil. A notable exception is the University of Sao Paulo (USP) in Brazil, which has the Russian Academy of Sciences (RAS) as its second biggest collaborator, after France's CNRS.

This unexpected connection appears to stem from a handful of influential Russian professors who are based at USP. One of the first to arrive, in 1992, was Dmitri Guitman, a physics professor from Siberia. Guitman has since trained many PhDs and postdocs from Russia and elsewhere. One of these, Maxim Dnovnikov, is back in Russia after three years at USP. While in Brazil, he continued to collaborate with his former Russian colleagues, reinforcing the USP–RAS relationship. "I didn't want to interrupt my previous collaborations, since post doc is a temporary job," says Dnovnikov. Another factor is the low salaries in Russia, which drive scientists to seek opportunities outside. However, with Brazil's economy flagging, he suggests such opportunities in the region might become rarer. **L.N. ■**

Rubio, a biotechnologist from the National Autonomous University of Mexico (UNAM; Mexico's strongest institution in the index), is one of the chief authors on a paper<sup>5</sup> on the genetic sequences of four tapeworm species.

Garcia Rubio also observes that the region is growing in independence: "Collaboration within Latin America is becoming more common than it was a few years ago." He worked on the tapeworm paper with colleagues from Latin America as well as from countries including the United Kingdom, United States and Switzerland, and says there are benefits to working with both types of partner. On one hand, local researchers easily understand each other's constraints, he says, such as small budgets, long delivery times for reagents and having more junior scientists on board. Researchers in rich countries have access to more resources, are better connected and have a better track record of publishing in high-impact journals, which he says benefits the collaborating scientists from poorer countries. **L.N. ■**



South American scientists help decipher the genetics of a pork tapeworm.

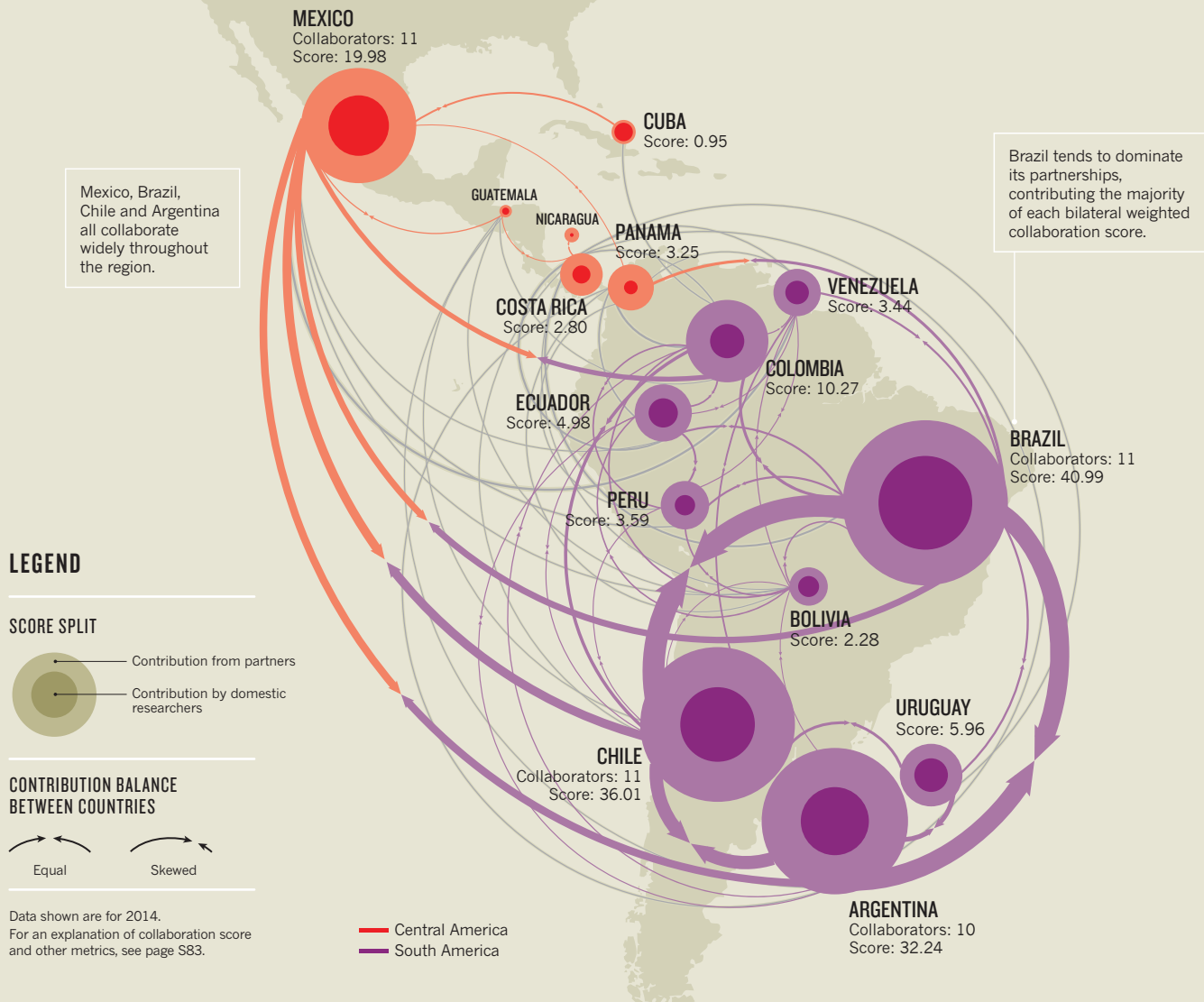
STEVE GSCHWEISSNER/SCIENCE PHOTO LIBRARY

1. Chennaoui Aoudjehane, H. *et al. Science* **338**, 785–788 (2012).
2. Johnson, K.P. *et al. Proc. R. Soc. B* **281**, 20132024.
3. Braga-Ribas, F. *et al. Nature* **508**, 72–75 (2014).
4. Castellón, J. *et al. MNRAS* **437**, 2607–2620 (2014).
5. Tsai, I. J. *et al. Nature* **496**: 57–63 (2013).

# TOWARDS SELF-SUFFICIENCY

Central and South American countries in 2014 have a stronger internal network than Africa, with the larger weighted collaboration scores shared across several countries rather than concentrated in one.

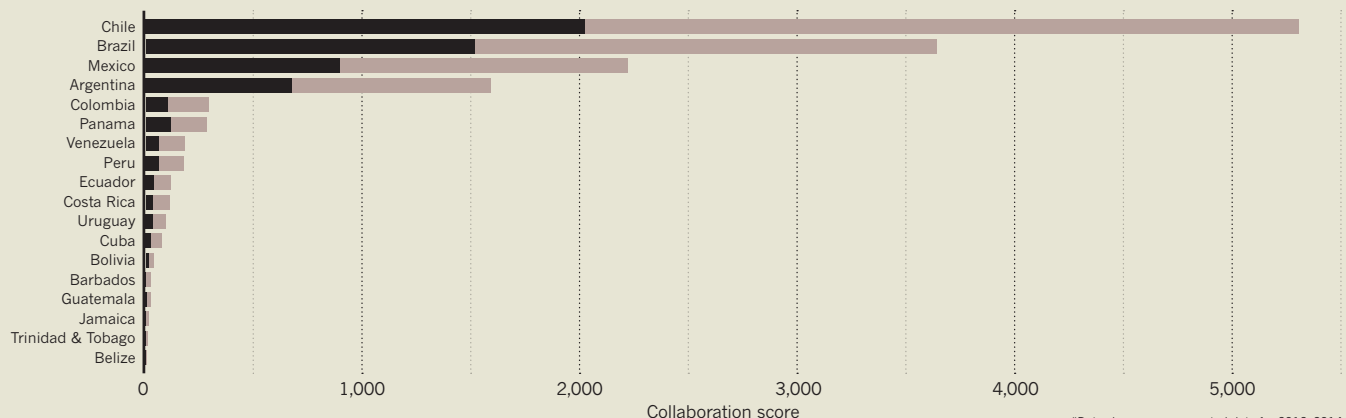
**NATURE.COM**  
To explore these graphics in more detail online, visit:  
[go.nature.com/VQAEi3](http://go.nature.com/VQAEi3)



## INTERNATIONAL LINKS

Using collaboration score removes the downweighting for astronomy. This gives Chile a distinct advantage over other countries in the region when it comes to international partnerships over the longer term.

Own contribution  
Partner country's contribution





Genetic therapy research at a laboratory shared by the Tianjin Medical University and the University of Oxford, which has many links with Chinese institutions.

# CHINA'S DIASPORA BRINGS IT HOME

*Links formed by mainland China's large scientific diaspora and its increasing output of high-quality research make it an emerging centre of international collaboration.*

After more than a decade of strong investment in research and higher education, China is becoming an important partner for the scientific powerhouses of North America and Europe, and a growing hub for international collaboration. Mapping the many Chinese international collaborations in the Nature Index (see 'China's global network') demonstrates the extent to which Chinese scientists have become innovative contributors to, and leaders of, many international scientific communities.

Like many nations, China's biggest collaborator in the Nature Index is the United States, the biggest index contributor overall, with a collaboration score more than five times that of its next strongest collaborating country, Germany (See 'China's closest links'). For China a contributing factor is the large number of Chinese researchers who have spent time in the global science superpower. A large diaspora of Chinese-heritage scientists around the world, particularly in the United States, have forged bonds between researchers in China and elsewhere. Indeed, more than 25% of scientists in the United States are from abroad and many of them are Chinese, notes Caroline Wagner, an associate professor who studies science and technology policy at the John Glenn College of Public Affairs, The Ohio State University in Columbus.

The Chinese government has made huge

efforts to incentivize a return to China by Chinese-born scientists. The Thousand Talent Program, operated by the Central Origination Department of the Communist Party is a prominent example of its efforts. "Many Chinese scholars have studied in the United States, and this contributes to international connections that continue after returning to China," Wagner says.

Yet the growth in US-China collaboration has occurred in the absence of a supportive political mood in the United States. "We should be mindful of potential threats to the development of collaboration," says Richard Suttmeier, professor of political science, emeritus, at the University of Oregon in Eugene.

"Concerns over national security and economic competitiveness, found in the United States, in China, and in other countries are probably not conducive to more robust patterns of collaboration," Suttmeier adds.

However, collaborations between China and European countries have been partly driven by governmental efforts on all sides to exploit common interest and mutual benefits in strategic research areas, such as energy, public health and sustainable urbanization.

Recently, for example, the European Union and China launched a new co-funding mechanism to support joint research and innovation activities. Each year, more than €100 million from the EU's Horizon 2020 programme will

be matched by at least 200 million yuan from Chinese programmes, for projects that involve European and Chinese participants. Before Horizon 2020, the EU ran the Framework Programme 7 from 2007 to 2013 in which China was the third largest international partner country, with 383 Chinese organizations participating in 274 collaborative research projects that garnered €35 million of funding from the EU.

Chinese policy is to encourage international cooperation in scientific research. Each of its government's three major agencies for science and technology — the Ministry of Science and Technology, the National Natural Science Foundation of China and the Chinese Academy of Sciences — has its own international department to promote such collaborations.

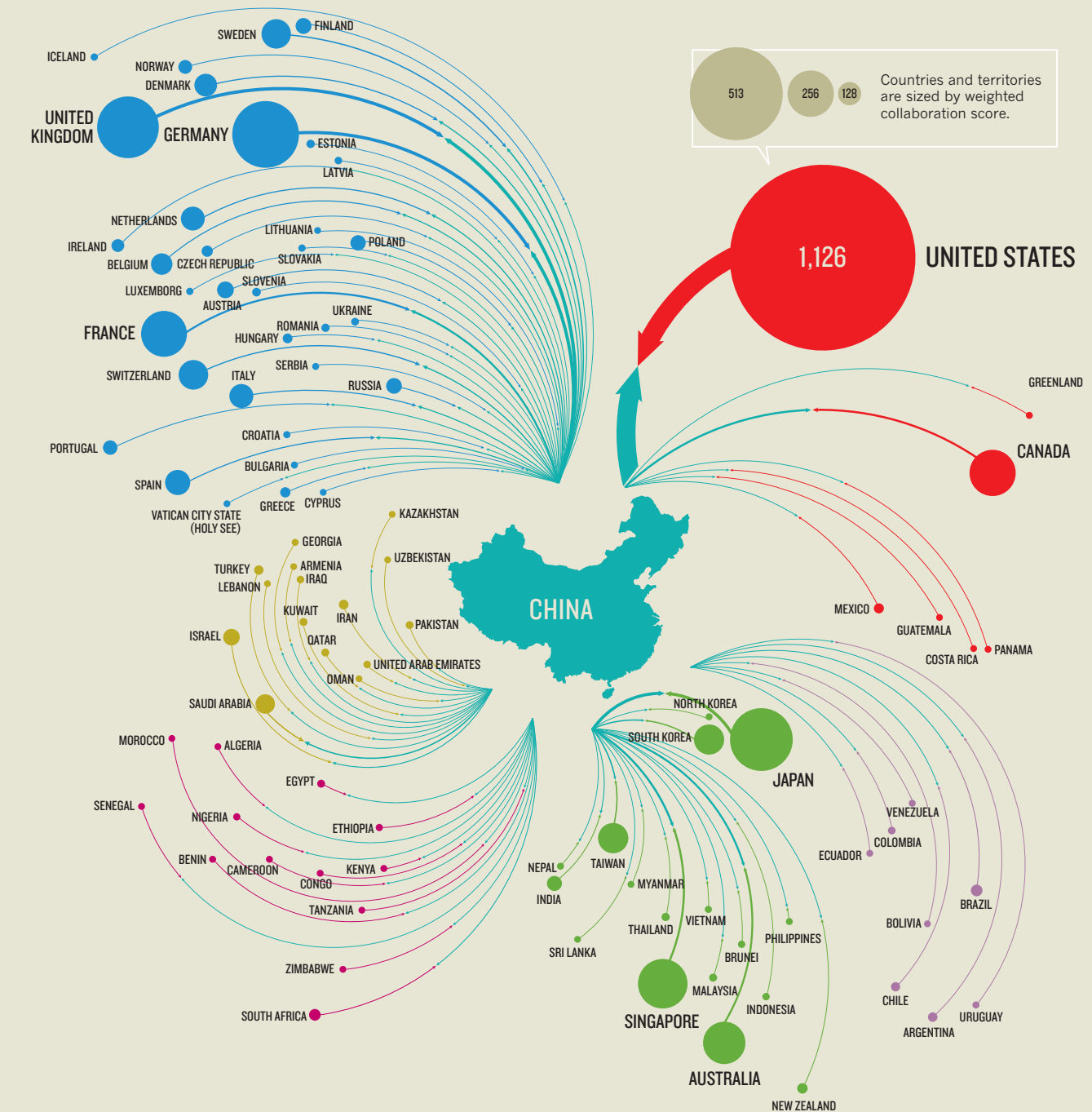
## SHARED PRIORITIES

In the Nature Index, the number of collaborations between China and the UK has been rising quickly (see 'Pulling its weight'). At the University of Oxford, a combination of shared academic priorities, policy support and researcher exchanges are helping to drive this acceleration. "There are almost 200 Chinese academics in Oxford, and more than 4,000 Oxford alumni around China," says Tao Dong, an immunologist at Oxford. Dong is also the Oxford-side founding director of a Centre for Translational Immunology (CTI) **PAGE S71 ►**



# CHINA'S GLOBAL NETWORK

In 2014, scientists from mainland China collaborated with counterparts in 94 other nations, making it one of the most connected countries in the Nature Index.



NATURE.COM  
To explore these graphics in more detail online, visit:  
[go.nature.com/VQAEi3](http://go.nature.com/VQAEi3)

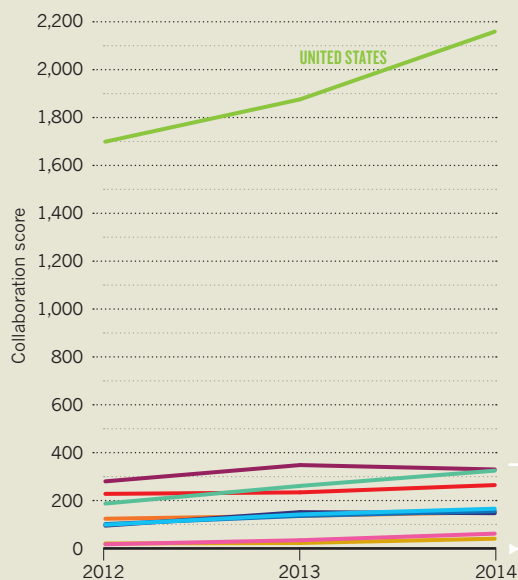


## CHINA'S CLOSEST LINKS

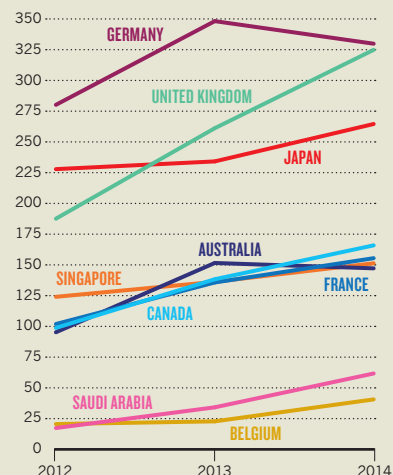
Tracking the ebb and flow of the country's most significant research relationships.

### CHINA'S TOP INTERNATIONAL COLLABORATIONS

Year on year, the growth in collaboration between the United States and China dominates the Nature Index as measured by the share of authorship from the two countries. Among other nations, the United Kingdom has seen particularly strong growth in collaboration.

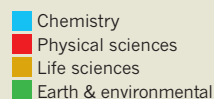


ENLARGED

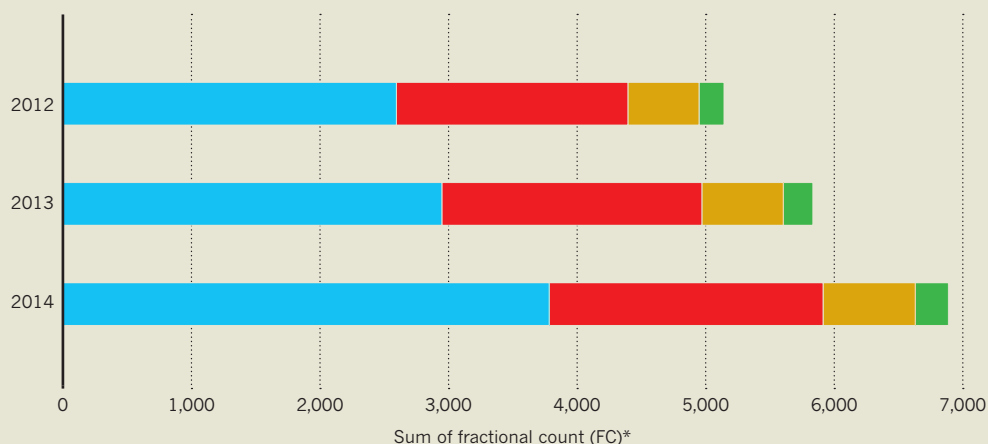


### GROWTH DRIVEN BY CHEMISTRY

China's output in the Nature Index has seen substantial growth year on year. A major driver of this appears to be output in chemistry.

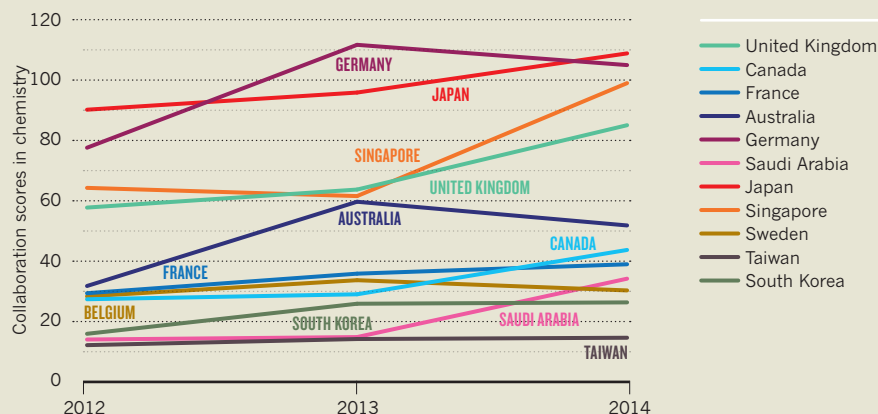


\*Subject areas can overlap. The sum of FCs derived from all the subject areas may therefore exceed the overall FC for China.



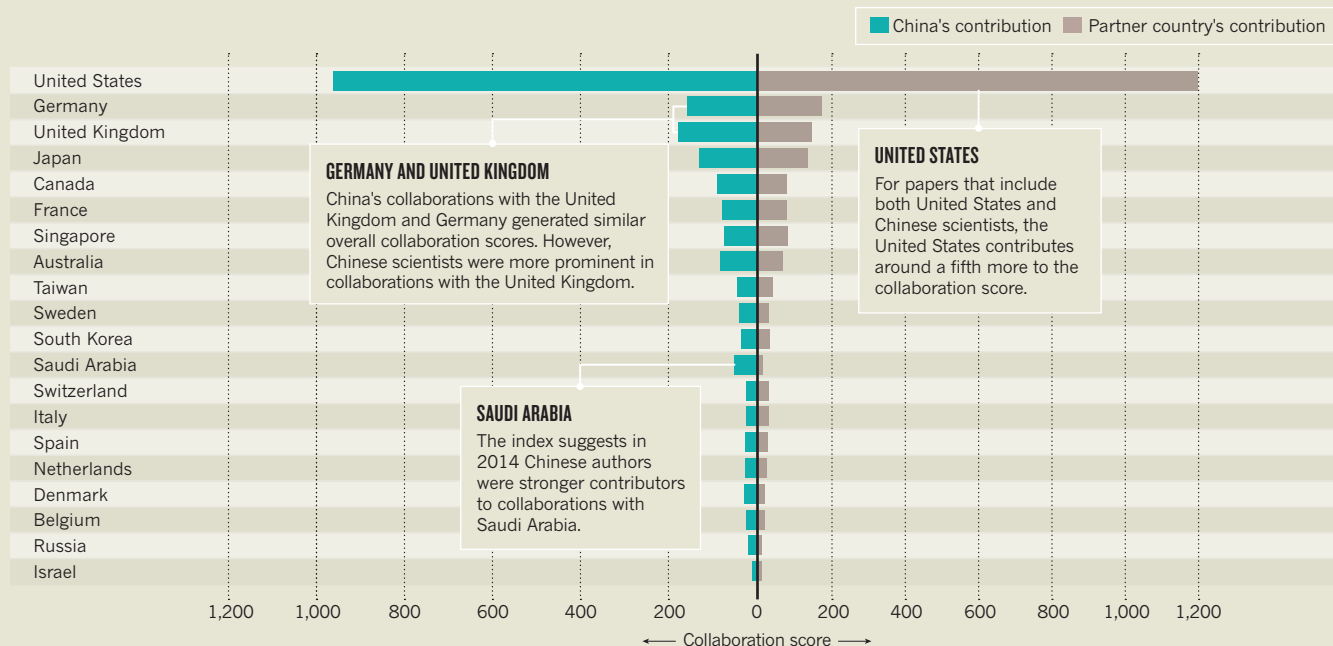
### CHEMICAL BONDS

Looking in detail at China's collaborations in the field of chemistry, major collaborators are the United States (not shown) followed by Germany and Japan. Collaboration with Singapore has undergone a notable increase, making it China's fourth-largest collaborator in this field.



## PULLING ITS WEIGHT

China is a major contributor to its most significant international collaborations in the Nature Index in 2014. This graph shows the proportion of the total collaboration score that can be attributed to authors from mainland China.



► at Oxford, which is a joint venture between the Chinese Academy of Medical Sciences and Oxford, as well as two other Chinese institutions. Dong believes that academic interest is the driving force for collaboration. “Research relevancy, mutual respect, and clear and upfront discussion on resources to facilitate research are essential.”

She adds that the support of key individuals in China, such as the Minister of Health, the mayor of Beijing, the president of the Chinese Academy of Medical Sciences, and the presidents and directors of universities and hospitals, has also been influential. “The CTI for example was endorsed as an area of collaborative research by the then Chinese Minister of Health during a visit of Oxford’s VC to China.”

### RESEARCH CHEMISTRY

Much of the increase in China’s output in the Nature Index since 2012 has been in chemistry (see ‘Growth driven by chemistry’). For example, there has been a dramatic rise in collaboration score between China and Singapore, with the partnership between the Hong Kong University of Science and Technology (HKUST) and the National University of Singapore (NUS) a notable example.

Two key players in this relationship are Benzhong Tang, in the Department of Chemistry at HKUST, who in 2001 discovered aggregation-induced emission, a novel photophysical phenomenon that enables light emission in the solid state, and Bin Liu, at the Department of Chemical and Biomolecular Engineering at

NUS, a bionanotechnology researcher investigating new materials. The collaboration has resulted in “more than 50 joint publications in high impact journals in the past four years”, says Liu. “The perfect matching of expertise and the nature of our complementary research has led to fruitful collaboration.”

“We work together not only to improve our own research, but also to help promote the whole community,” Liu adds. “There is now a quite large community of aggregation-induced-emission researchers.”



Luminogens prepared by Benzhong Tang’s team.

### PHYSICAL ATTRACTION

The United States aside, in the physical sciences China’s international collaborations in the index are dominated by Germany. The relationship between the University of Science and Technology of China (USTC) and the University of Stuttgart is particularly prominent.

These collaborations arise from practical, research-driven necessity, says Nan Zhao, a theoretical physicist at the Quantum Optics & Quantum Information group, Beijing Computational Science Research Center, who has coauthored papers with the scientists from USTC and Stuttgart.

“Some experiments simply cannot be done in China,” Zhao says. For instance, Germany’s long history in the physical sciences means it has a capacity for sample preparation and manufacturing of experimental equipment that China lacks, says Zhao. German scientists, such as Zhao’s collaborator Jörg Wrachtrup at Stuttgart, also provide valuable mentorship to their relatively inexperienced Chinese partners.

“Generally, we are ‘students’, and cannot start and lead an innovative research field yet,” says Zhao. Now based in Beijing, he has spent time at the University of Stuttgart, helping to build mutual understanding and trust with his collaborators.

“Innovative ideas are like sparks by chance, which cannot be planned or organized into projects. They can only be captured by deep and regular academic communication,” he says, “What the governments should not do is to set up barriers.” P.T. ■

BENZHONG TANG

# INDUSTRIAL-STRENGTH BONDS

*Though industry collaboration with academia often happens between neighbours, communication and shared goals outweigh proximity in driving success.*

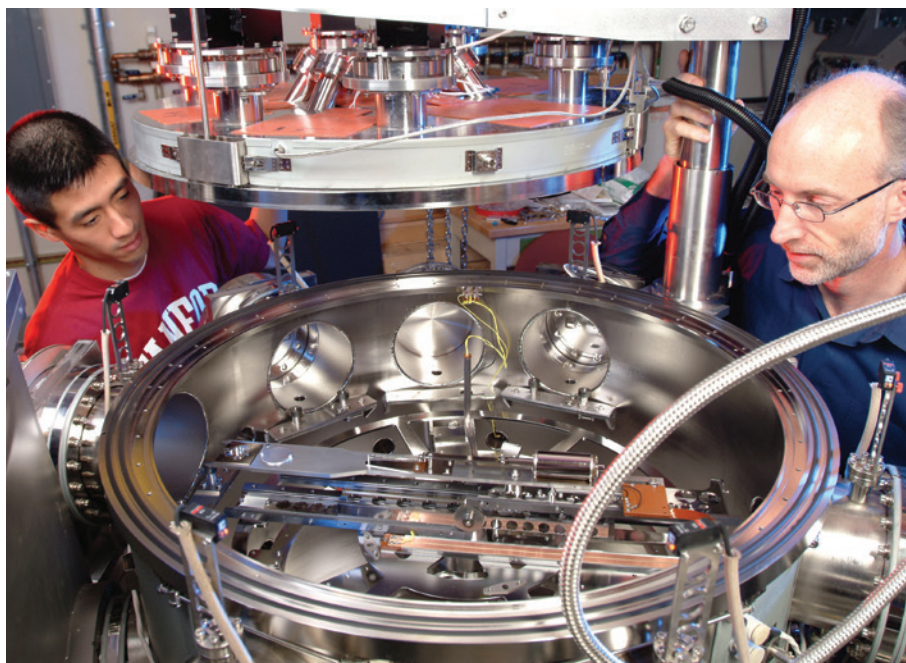
IBM RESEARCH

Although academic institutions dominate the natural-science focused Nature Index, there is notable contribution from industry, which in 2014 represented nearly a quarter of institutions in the index, contributing to nearly 7% of articles. Analysis of collaborations in the index from 2012 to 2014 reveals several global hotspots in East Asia, Northwest Europe and the United States where there are multiple industry-academia collaborations (see 'Industrious world'). And although there are several factors behind this geographic clustering, ease of communication is eroding many of them.

"The barriers to global collaborations are very low right now," says Edmund Graziani, head of bioconjugation chemistry and natural products discovery at Pfizer in Groton, Connecticut, who frequently collaborates with academics. New York-based Pfizer has more than half of its collaborations with institutions outside the United States (see 'Pfizer's partners') but Graziani says this is nothing out of the ordinary for a multi-national company. Pfizer is experienced at running long-distance projects, plus the company has several R&D centres across international locations that work together. "We don't typically see that as a barrier to collaboration," he says.

Indeed, it seems that location has little effect on the success of a collaboration, according to a 2010 study<sup>1</sup> by researchers from MIT, which interviewed 25 research-intensive multinational companies. Study author, Julio Pertuzé, now at the Pontificia Universidad Católica in Chile, says that although geography has historically been considered integral to industry-university links, this is less true for collaborations between large firms and prestigious universities — those most often found in the Nature Index. There is a trend for clustering of businesses around top universities, though no established pattern for how these clusters start. "It is a little bit like the chicken and the egg in terms of how industry-university relationships evolve," Pertuzé says. Such technology clusters can include university spin-off companies, larger corporate research centres and collaborative industry-university research initiatives, all attracted by the expertise the university can offer, and the networking opportunities within the cluster itself.

In Japan, Osaka is the second largest industry-academia collaboration cluster after Tokyo. The city is home to several renowned universities including Osaka University as well as many electronics and pharmaceutical firms,



IBM researcher Stuart Parkin (right) with Stanford graduate student Roger Wang inspecting part of the thin-film deposition system that underpins spintronics technology at IBM's Almaden Research Center.

such as Takeda Pharmaceutical Company, Sumitomo Electric Industries and Shionogi & Co. Ichio Shimada, a physical chemist in the University of Tokyo, has worked with Shionogi for five years. Shimada says the company contacted him because of its interest in his published work and conference presentations on cell membrane protein receptors and their molecular interactions.

**"IT SEEMS THAT LOCATION HAS LITTLE EFFECT ON THE SUCCESS OF A COLLABORATION."**

Shionogi proposed a collaboration that closely paralleled his work under which experiments were carried out in Shimada's lab in Tokyo with company researchers who also provided reagents. The collaboration resulted in one paper<sup>2</sup> in the Nature Index in 2012, on monitoring the interactions between drug molecules and protein receptors. Shimada says he met Shionogi management to discuss project progress every four or five months, making the 500-km journey between Osaka and Tokyo in only three and a half hours on a bullet train.

"They collaborate where there is what they need," he says, of Shionogi. "Location is not a big issue. Japan is not as large as the US."

In France, university-industry collaboration often happens through 'mixed' laboratories. One such physics lab, Unite Mixte de Physique CNRS/Thales (UMPhy), involves electronic systems company, Thales Group (previously Thomson-CSF), which is France's leading company for academic collaborations in the Index (see 'Parisian firms'). The Paris-based lab was created in 1995 by France's National Scientific Research Centre (CNRS), Thales and the Université Paris-Sud. UMPhy grew out of a longstanding magnetic materials collaboration between Nobel Prize winner Albert Fert of Paris-Sud and Thales Central Research Lab. It is considered the birthplace of spintronics — an emerging branch of electronics that manipulates the spin of an electron in addition to its charge. Frédéric Petroff, deputy director of UMPhy, says that the lab "strongly favours interactions with Thales researchers and engineers", having been set up to stimulate what he calls "the coffee machine effect" — where new ideas are conceived by colleagues chatting while taking a short break.

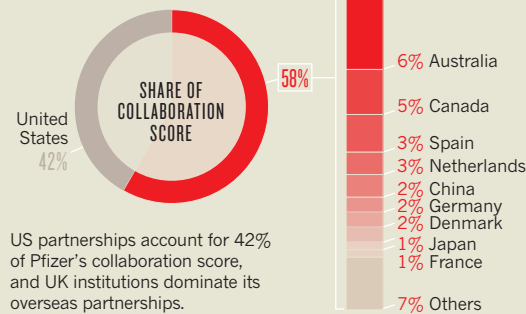
Javier Villegas, a physicist in UMPhy, has two collaborative publications in the index, on superconducting and magnetic **PAGE S79 ►**

# INDUSTRIOUS WORLD

Companies publishing in the Nature Index cluster in three regions: East Asia, Northwest Europe and the United States. Tokyo is the leading city by collaboration score, followed by Basel in Switzerland, then London, although most of the rest of the top 20 is dominated by US cities.

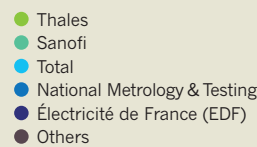
## PFIZER'S PARTNERS

Pharmaceutical firm Pfizer is based in the United States, but only 30% of its partnerships in the index are within the same country.



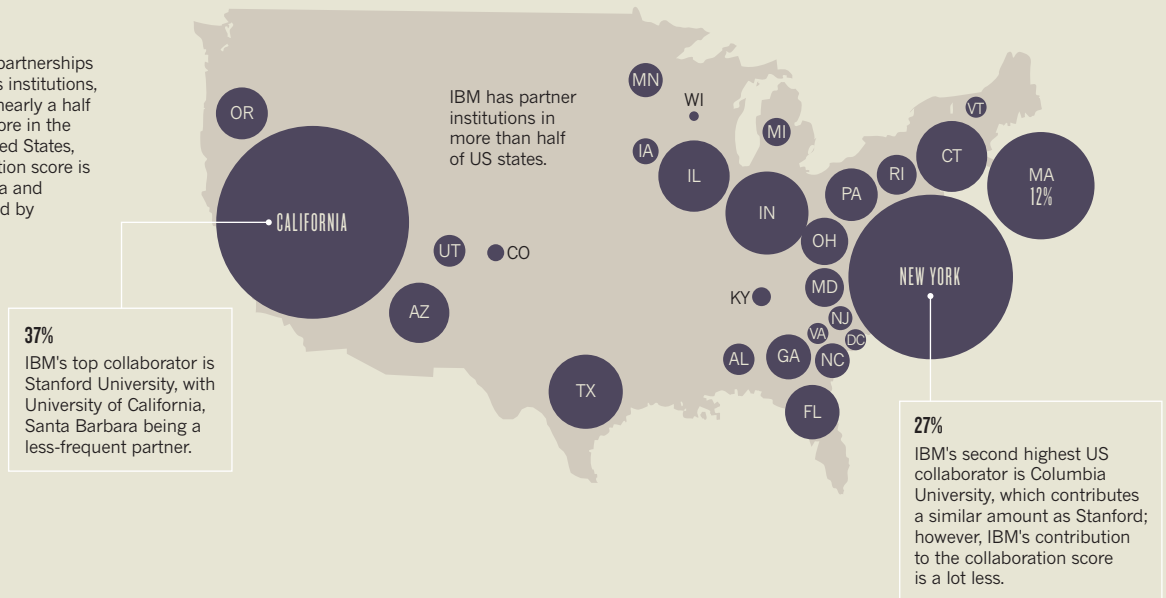
## PARISIAN FIRMS

Paris is the seventh industry-heavy city in the index. The top 10 companies with headquarters in Paris account for nearly 90% of its collaboration score.



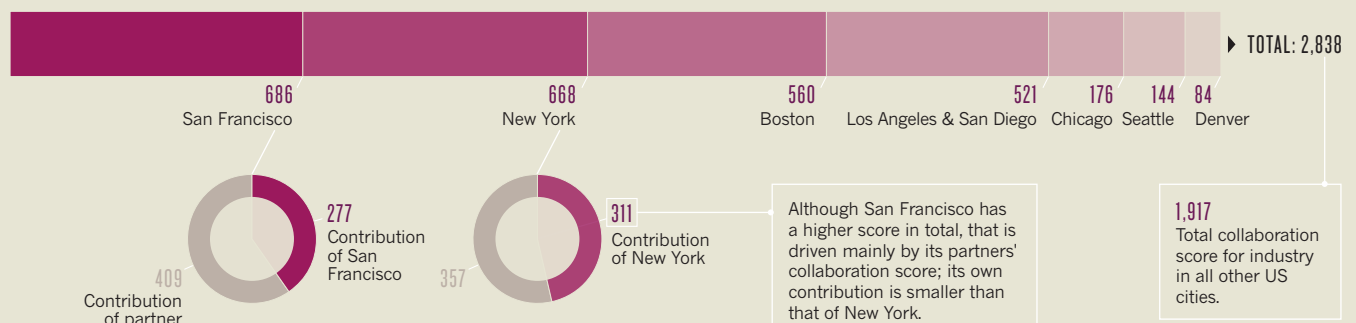
## IBM AT HOME

Only a third of IBM's partnerships are with United States institutions, but they account for nearly a half of its collaboration score in the index. Within the United States, most of the collaboration score is generated in California and New York.



## US HUBS

Four large coastal cities dominate the United States in terms of industry output in the index.

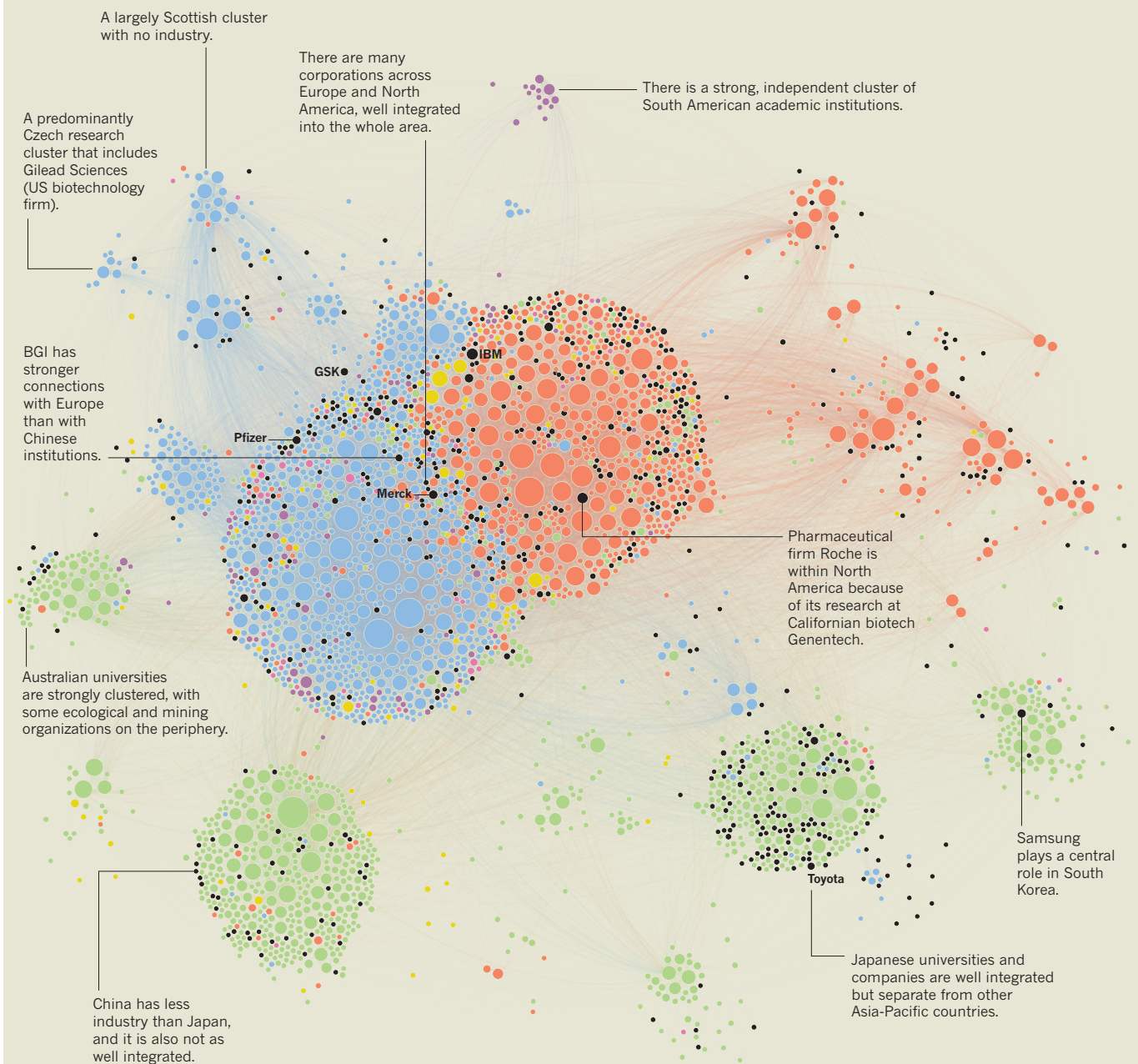


Data shown are aggregated for 2012-2014.  
For an explanation of collaboration score and other metrics, see page S83.



## INDUSTRIAL MIXING

All institutions in the index are shown here, sized by fractional count (FC) and coloured by geographical region. A pair of institutions is connected when they collaborate on a paper — the more papers, the stronger the connection and the thicker the connecting line. Clusters of institutions are determined algorithmically and depend on the size of each institution and the strength of its various connections (analogous to many charged balls that repel each other, connected by springs of different strengths). The diagram shown emphasizes certain country clusters, with industrial institutions picked out in black. The distribution of corporations within the clusters reveals information about their relationships with other non-commercial (mostly academic) institutions.



● Africa ● Asia Pacific ● North America ● Europe ● South America ● West Asia ● Industry

Data shown are for 2014.

For an explanation of collaboration score and other metrics, see page S83.

**NATURE.COM**

To explore the graphic in more detail online, visit:

[go.nature.com/i2LVfb](http://go.nature.com/i2LVfb)

► materials<sup>3,4</sup>. Villegas says interaction with the 22% of industrial researchers in the lab is helpful for the academics. “It allows me to identify and imagine concrete ways in which my more basic, fundamental research could address current technological challenges,” he says. “Conversely, Thales researchers can find novel technological applications based on my research findings. This symbiosis is extremely profitable for everyone.” There are other benefits of industry-academia tie-ups, says Marko Erman, chief technology officer at Thales. “The tax reduction mechanism in France (the so-called *Crédit d’Impôt Recherche*) makes such cooperation financially attractive for both partners.” The benefits of working in such close proximity are embraced by Thales, which has four other joint laboratories in the United Kingdom, the Netherlands, Singapore and Canada.

### PUSH AND PULL

The Nature Index also contains many collaborations between companies that have based a research centre near a major university. New York-based IBM is the company with the second highest collaboration score in the index, after China’s BGI. In the heart of Silicon Valley, IBM has its San Jose-based Almaden Research Center, which works closely with nearby Stanford University — its most frequent collaborator, with 25 jointly authored papers in the Index (See ‘US hubs’). Former Stanford electrical engineering student and current IBM Almaden researcher, Timothy Phung, has experienced both sides of such collaborations. His 2013 index paper<sup>5</sup> was the basis of his PhD work, and part of a continuing ten-year collaboration involving his Stanford supervisor, electrical engineer, James Harris, Stanford physicist, Shouheng Zhang and the IBM team of Stuart Parkin. The group, known as the IBM–Stanford Spintronic Science and Applications Center, is developing magnetic materials for a new type of high density solid-state computer memory known as domain-wall or ‘racetrack’ memory. “Stanford is a very entrepreneurial place, and collaborations of this type with companies are very much welcomed,” says Phung.

IBM Almaden lab director, Jeff Welser, says setting up near university and innovation centres “has certainly contributed to our ability to collaborate effectively”. He says both Stanford and IBM staff attend the same symposia — leading perhaps to a ‘coffee break’, rather than coffee machine effect. “It is also very easy for our researchers to spend a day or more a week at the university[ ... ] facilitating improved knowledge transfer, more opportunities for co-invention, and joint facility usage,” says Welser.

Sometimes, a collaborative push might come from academia. Harvard Medical School microbiologist Gerald Pier started his collaboration with French pharmaceutical firm Sanofi in 2009, after consultants helped him search for a commercial partner. As well as his academic post, Pier has a start-up company, Alopexx

BGI is a genome sequencing centre based in Shenzhen, southeast China. It was founded in 1999 as the Beijing Genomics Institute, a private and independent research organization, but has since developed as a corporate entity. Its many links have earned it a collaboration score of 322, the highest for any company in the Nature Index and 23% more than second placed IBM. BGI is one of China’s largest commercial providers of genomic sequencing and analysis services.

Index data show that BGI’s most frequent academic partner is the University of Copenhagen (see ‘Academic alliances’). This unexpected association began even before BGI was founded. Yang Huanming, one of the founders of BGI, got his PhD from Copenhagen in 1988. “The friendship between BGI and the University of Copenhagen has a strong basis in mutual understanding and trust,” says Yang. Wang Jun, the former chief executive of BGI, is also a professor in Copenhagen’s department of biology. Yang says, there are more than 20 researchers at BGI who got their PhD

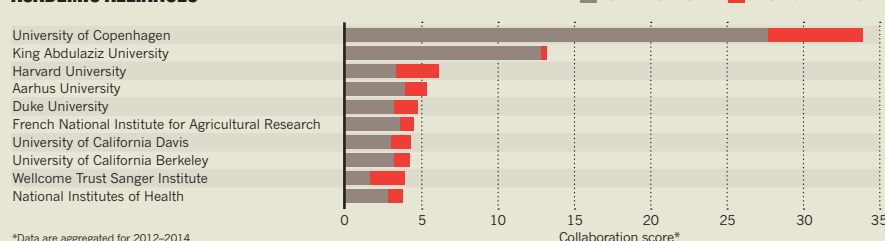
in Denmark, mostly from the University of Copenhagen.

Strengthening this relationship, BGI founded its European branch in Denmark in 2010, adding a research centre in Copenhagen Bio Science Park a year later. Researchers from the university provide samples and ideas, and BGI performs the genomic sequencing and analysis — notable examples include meta-genomic sequencing of human gut bacteria<sup>7</sup> and helping piece together ancient DNA<sup>8</sup>.

BGI and Copenhagen also apply jointly for funds from the European Union and from the National Natural Science Foundation of China.

Because these projects are highly collaborative, BGI typically gets a very low fractional count in the Nature Index for each paper. Yang insists that BGI’s role is nonetheless valuable. “Mega science needs vast scientific collaborations,” he says. “It should be regarded as an important contribution to the research work to organize so many institutions and researchers into collaborations.” **P.T. ■**

### ACADEMIC ALLIANCES



Pharmaceuticals, as a side project, and it was through Alopexx that he collaborated with Sanofi on an antibody treatment for serious bacterial infections as an alternative to antibiotics.

The partnership led to a 2013 index publication<sup>6</sup>, and in 2014 the collaboration successfully ended with Sanofi licensing the technology. Pier says the long distance between Boston, Massachusetts, and Toulouse, France, posed some problems such as time-zone differences, but essentially the issues were the same as with any collaboration, such as working out how to establish protocols and share data.

The greater Boston area, home to several leading universities and teaching hospitals as well as global medical technology firms, has the third largest industry-academia collaboration score in the United States. Pier says the local environment encourages academic interest in commercializing technologies. But he also suggests another reason why such collaborations seem to cluster in regions like Boston: he has seen a rise in the number of “two-career couples”, who will be looking to work at different organizations in the same location. Pier admits, however, that this effect is difficult to quantify.

So does location matter? Julio Pertuzé says his work showed that it is communication, not distance, that matters. “Proximity is only important if firms act upon the opportunity for increased face-to-face interactions,” he says. But he adds that, in many collaborations, “there are other options for increasing these interactions that have little to do with geographic distance.” Pertuzé says these include relocation of industry personnel to work in a partner’s lab, or facilitating the exchange and travel of researchers between sites. And while location plays a part, ultimately “it is a question of the specific technology or innovative approach that a given academic brings to bear,” Graziani explains, “the quality of science far outweighs location.” **R.B. ■**

1. Pertuzé, J. A. *et al. MITSloan Manage. Rev.* **51**, 83 (2010).
2. Kofuku, Y. *et al. Nature. Commun.* **3**, 1045 (2012).
3. Visani, C. *et al. Nature. Phys.* **8**, 539 (2012).
4. Trastoy, J. *et al. Nature. Nanotech.* **9**, 710 (2014).
5. Pushp, A. *et al. Nature. Phys.* **9**, 505 (2013).
6. Cywes-Bentley, C. *et al. Proc. Natl Acad. Sci. USA* **110**, E2209 (2013).
7. Qin, J. *et al. Nature* **490**, 55 (2012).
8. Carpenter, M. L. *et al. AJHG* **93**, 852 (2013).



# OPENING BORDERS AND BARRIERS

*Collaboration may result in higher impact science, but are government initiatives the best way to promote such international and interdisciplinary connections?*

KAVLI INSTITUTE

An American physicist, a Japanese mathematician and a German cosmologist walk into a lab; what do you get? Based on recent outcomes, you'll get ground-breaking science. And lately, governments have begun paying heed to evidence<sup>1</sup> that suggests international, multidisciplinary collaborations such as these will yield high-impact results.

Policymakers from diverse countries, including China, Japan, Australia, Chile and Germany, have sought to foster excellent science and technological innovation — and reap the associated economic benefits — by promoting collaboration across borders and disciplines, and setting up specialist centres with the necessary resources (see 'Conduits to collaboration').

But does this kind of top-down approach really increase the number of collaborations beyond what may have occurred organically? And, more importantly, is it producing the best possible research? Here, the answers aren't so definitive. Nature Index data show that some specialist centres are much more collaborative than others, and that their collaboration scores range from several hundred down to a handful.

## TEA AND TEAMWORK

The scenario described in the opening line is what you may find any weekday afternoon in the atrium of the Kavli Institute for the Physics and Mathematics of the Universe (IPMU) on the campus of the University of Tokyo, Japan. The institute has a mandatory staff tea time during which ideas and avenues for potential collaboration are shared.

Tea time is a custom that Kavli IPMU director and particle physicist, Hitoshi Murayama, brought back to Japan from the United States after witnessing its benefits in labs he had worked in. He was so convinced of its value in promoting collaboration among researchers that he put it in the initial proposal for the institute and battled with the funders to have it included. "I don't think taxpayers anywhere like to pay for tea, but I knew that it was worth fighting for," he says.

The Kavli IPMU is a product of the Japanese Government's World Premier International Research Center Initiative (WPI), launched in 2007 with the aim of creating "globally visible research centres". With more than 350 collaborating partners in index papers in 2014, Kavli IPMU is the most prolific of the nine WPIs (see 'Teams of all sizes').

These WPIs are part of a growing collection of initiatives established by governments



Tea time at Kavli Institute allows for an organized and informal exchange of collaborative ideas.

around the world with similar goals. Prominent examples include the ARC Centres of Excellence in Australia, the Millennium Science Initiative in Chile, Germany's Clusters of Excellence, and, most recently, the Collaborative Innovation Centres in China.

**"I DON'T THINK  
TAXPAYERS LIKE PAYING  
FOR TEA, BUT IT WAS  
WORTH FIGHTING FOR."**

The common goal underpinning these schemes is to gather a critical mass of leading researchers to work together on cutting-edge projects that will potentially have an impact on the world stage. Each of the initiatives mentioned has an explicit mandate to increase research collaboration, both domestically and internationally, although there is less emphasis on domestic partnerships in the Japanese and Chilean models. The research bodies created under these schemes are given significant autonomy to determine their own research priorities, and funding is granted for substantial periods, a decade or more in the case of WPIs, to enable longer-term research projects to be planned with confidence.

For the WPIs, guidelines specify that at least 30% of the researchers at any institute must be from overseas, and English is to be the primary

language. Still less than a decade old, Kavli IPMU has ticked off all requirements, and is achieving impressive research collaboration.

Kavli's core business, astrophysics and astronomy, both demand enormous collaborative efforts. Along with the geosciences and high-energy physics — where a single paper can have as many as 3,000 authors — these are the realms of science most obviously suited to such collaborative centres. In Europe, for example, the TUM Cluster of Excellence for the Origin and Structure of the Universe has the highest collaboration score among Germany's Clusters of Excellence.

Away from physics, astronomy and geosciences, the Munich Cluster for Systems Neurology (SyNergy) is the most collaborative of all the government-backed collaborative research institutions included in the Nature Index analysis. In 2014, it shared authorship with well over 100 separate research bodies — the majority fellow German institutions.

At the other end of the scale, China's relatively young Collaborative Innovation Centers — the first were established in 2012 — have fewer partners, particularly when it comes to international institutions. But it seems likely that numbers will increase before long.

## RESEARCH POLICY RISK

Some observers caution that such government-directed research initiatives remain focused on generating the highest quality science at the expense of other policy goals.

## CONDUITS TO COLLABORATION

Governments around the world have established special centres with the ambition of fostering high quality research through collaboration. This graphic shows up to 10 of the top centres per country in the Nature Index 2014.

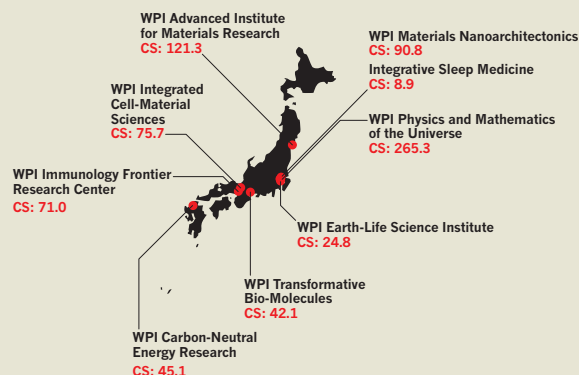
### CHINA

China's first Collaborative Innovation Centers (CICs) were established in 2012. There are 30 represented in the index in 2014.



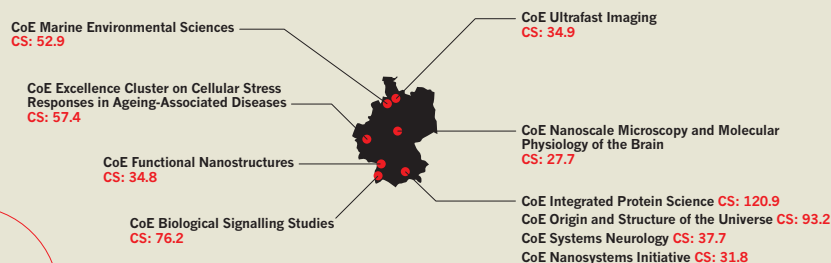
### JAPAN

Japan's World Premier International Research Center Initiative (WPI) aims to create "globally visible research centres". Nine appear in the 2014 index.



### GERMANY

Germany's DFG funds 43 Clusters of Excellence (CoE) currently, of which 27 appeared in the Nature Index in 2014.



Countries here are shown at accurate relative sizes.

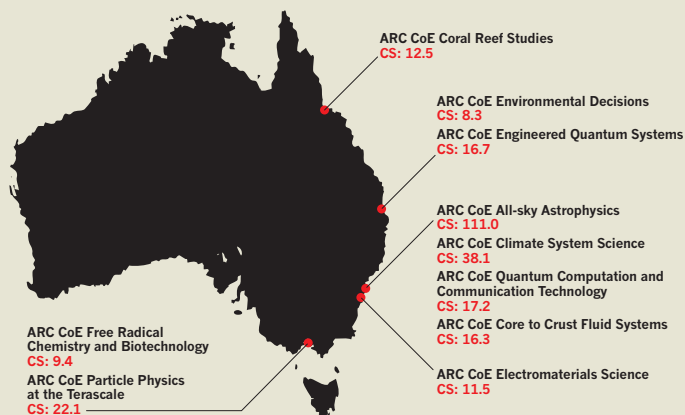
### CHILE

Four World Bank-backed Millennium Institutes in Chile are represented in the 2014 index.



### AUSTRALIA

In 2014, 22 of Australia's 30 ARC Centres of Excellence (CoE) had a collaboration score in the Nature Index.



CS = Collaboration Score.

Data shown are for 2014.

For an explanation of collaboration score and other metrics, see page S83.

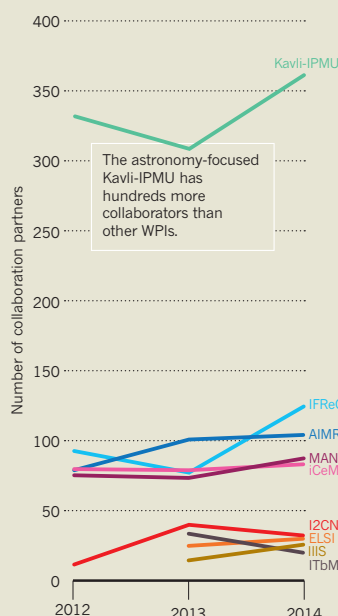


## TEAMS OF ALL SIZES

Japan's World Premier Institutes (WPI) illustrate the variability of collaborative patterns across disciplines.

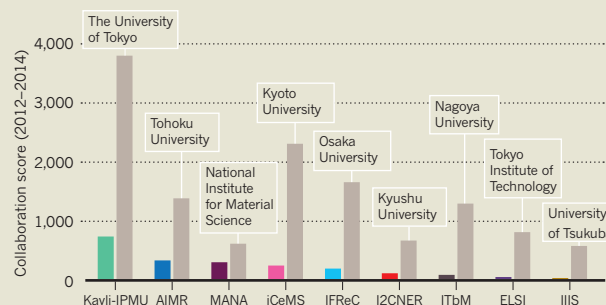
## STARS ABOVE

The number of collaborative partners in index papers varies significantly between Japan's WPI Institutes.



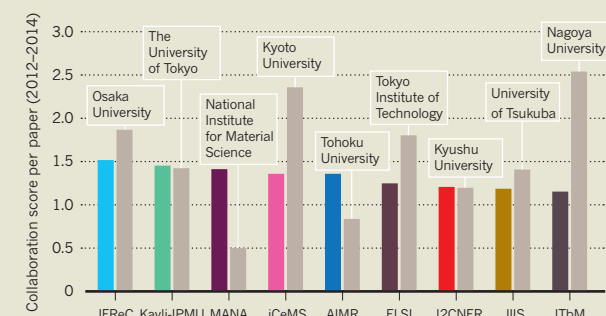
## UPS AND DOWNS

Comparing Japan's WPIs to their parent institutions by collaboration score over 2012–2014.



## ALL RELATIVE

Looking at collaboration score per collaborative paper, some WPIs perform better than their host institutions.



● WPI Physics and Mathematics of the Universe (Kavli-IPMU)  
● WPI Immunology Frontier Research Center (IFReC)  
● WPI Advanced Institute for Materials Research (AIMR)  
● WPI Materials Nanoarchitectonics (MANA)  
● WPI Integrated Cell-Material Sciences (iCeMS)  
● WPI Carbon-Neutral Energy Research (I2CNER)

● WPI Earth-Life Science Institute (ELSI)  
● WPI Integrative Sleep Medicine (IIIS)  
● WPI Transformative Bio-Molecules (ITbM)

Caroline Wagner, an Ohio State University expert in the links between science, policy, society, and innovation, advises that a distinction should be made between the kind of collaborative activity organized by scientists themselves — where the research determines the organizational structure of the teams — and the government-driven collaborations that have different objectives. “These can be about a lot of other things in addition to science, for example, national prestige, industrial competitiveness and encouraging intellectual property development.”

She says researchers should resist policy initiatives that get in the way of achieving the highest scientific standards — pointing to European Union rules dictating that participants from at least three different EU states must be involved in any successful proposals — a policy designed to lift the performance of those states with weaker science records.

“Adding these policy goals into the science almost always reduces the quality of the science, but policymakers are willing to buy down the quality of the science in order to reach other kinds of social or political goals.”

Les Field, secretary of policy with the Australian Academy of Science, shares Wagner's observations about the potential pitfalls of mixing science and policy, but says the success he has seen achieved by the best Australian centres of excellence and their international equivalents, producing great science and establishing collaborations at the highest levels, should be

acknowledged and encouraged by governments. “It's important to recognize the benefits that you get out of the continuity of some very high-powered teams which can contribute very significantly to the reputation and economic progress of the country.

Indeed, there is good science being produced by collaborative institutions included in the index.

**“A CRITICAL MASS MEANS THAT INTERNATIONAL LINKAGES FORM FAR MORE EASILY.”**

At the Kavli IPMU, for example, collaboration with researchers from a range of Japanese institutions and from Duke University in the United States, led to the recent discovery of a gravitational lens in space created by a hitherto unknown galaxy which explained the superluminescence of an unusual supernova that had been puzzling astronomers<sup>2</sup>.

Elsewhere, research from Australia's ARC Centre of Excellence for Climate System Science, in collaboration with researchers from France's Laboratoire de Meteorologie Dynamique, has significantly increased the sensitivity ascribed to the effects of increased carbon in the earth's atmosphere in climate modelling<sup>3</sup>.

Director of the ARC centre, Andrew Pitman, says the rapidly increasing complexity of climate modelling over the last decade, requiring computer coding with as many as one million lines, means that the field has become dependent on collaboration facilitated by centres such as his. “No one single country can build the climate models and maintain international competitiveness in modelling systems so we collaborate with the big groups in the US, Germany and Britain.”

Pitman says that the rules applied to work at the centres of excellence, which force collaboration, allow them to more easily overcome the structural biases that exist against sharing research credits with other institutions.

Meanwhile, he says, having a critical mass of researchers working together with good funding over a long period enables necessary software experts and technicians to be employed and widely shared, and for younger scientists to receive training.

Bringing researchers into such centres also allows them to travel and work with peers around the world — establishing the links necessary for future collaborations — Pitman says. “It may just occur naturally but I think having the critical mass means that the international linkages form far more easily.” **A.M. ■**

1. Narin, F. *et al. NASIF* **21**: 2–3 (1991)
2. Quimby, R. *et al. Science* **344**: 396–399 (2014)
3. Sherwood, S. *et al. Nature* **505**: 37–42 (2014)

# A GUIDE TO THE NATURE INDEX

*A description of the terminology and methodology used in this supplement, and a guide to the functionality available free online at [natureindex.com](http://natureindex.com).*

The Nature Index is a database of author affiliations and institutional relationships. The index tracks contributions to articles published in a group of highly selective science journals, chosen by an independent group of active researchers.

The Nature Index provides absolute counts of publication productivity at the institutional and national level, and as such is one indicator of global high-quality research output.

Data in the Nature Index are updated monthly, with the most recent 12 months of data made available under a Creative Commons licence at [natureindex.com](http://natureindex.com).

The database is compiled by Nature Publishing Group (NPG) in collaboration with Digital Science.

The list of journals tracked by the Nature Index is under review, and from 2016 will be extended to include the clinical sciences.

## NATURE INDEX METRICS

There are three measures provided by the Nature Index to track affiliation data. The simplest is the **article count** (AC). A country or institution is given an AC of 1 for each article that has at least one author from that country or institution. This is the case whether an article has one or a hundred authors, and it means that the same article can contribute to the AC of multiple countries or institutions.

To get a better sense of a country or institution's contribution to an article, and to remove the possibility of counting articles more than once, the Nature Index uses the **fractional count** (FC), which takes into account the relative contribution of each author to an article. The total FC available per paper is 1, which is shared between all authors under the assumption that each contributed equally. For instance, a paper with 10 authors means that each author receives an FC of 0.1. For authors who have joint affiliations, the individual FC is then split equally between each affiliation.

The third measure used is the **weighted fractional count** (WFC), which applies a weighting to the FC to adjust for the overrepresentation of papers in astronomy and astrophysics. The four journals in these disciplines publish about 50% of all papers in international journals in this field — approximately five times the equivalent percentage for other fields. Therefore, although the data for astronomy and astrophysics are compiled in the same way as for all other disciplines, articles from these journals are assigned one-fifth the weight

Users of [natureindex.com](http://natureindex.com) can search for specific institutions or countries and generate their own reports, ordered by article count (AC), fractional count (FC) or weighted fractional count (WFC).

Each query will return a profile page that lists the country or institution's recent research outputs, from which it is possible to drill down for more information. For example, articles can be displayed by journal, and then by article title. As in the supplement, research outputs are organized by subject area. The profile page also lists the institution or country's top collaborators, as well as its relationship with other research organizations.

of other articles (i.e. the FC is multiplied by 0.2 to derive WFC).

The total FC or WFC for an institution is calculated by summing the FC or WFC for individual authors.

The process is similar for countries, although complicated by the fact that some institutions have overseas labs that will be counted towards their host country totals. What's more, there is great variability in the way authors present their affiliations. Every effort is made to count affiliations consistently, with a background of reasonable assumptions.

For more information on how the affiliation information is processed and counted, please

see the FAQ at [natureindex.com](http://natureindex.com).

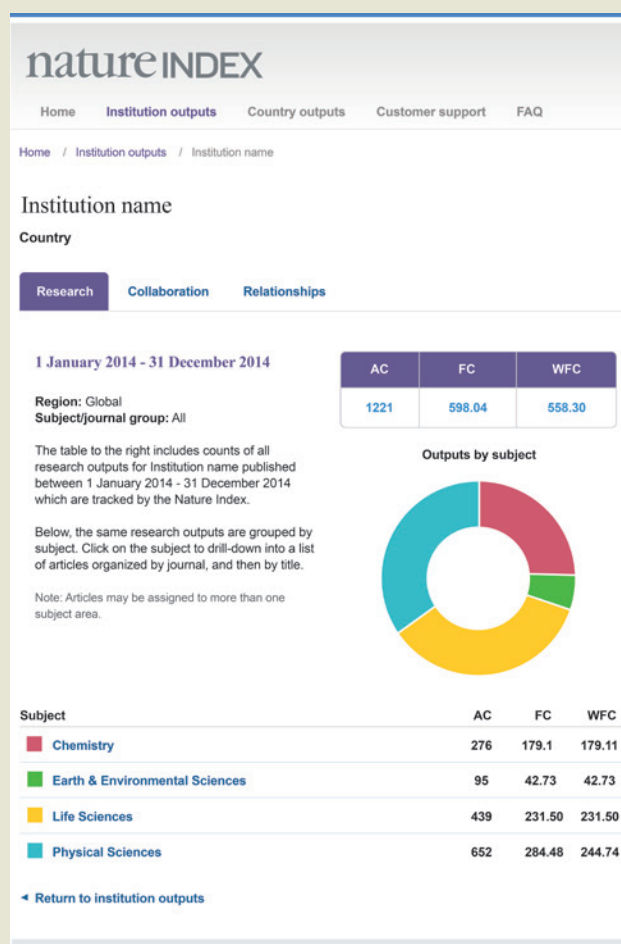
## THE SUPPLEMENT

*Nature Index 2015 Collaborations* is based on a snapshot of data from [natureindex.com](http://natureindex.com), covering articles published between 1 January and 31 December 2014.

Most calculations within the supplement use either the **weighted** or **unweighted collaboration score** for institutions or countries. This is derived by summing either the WFC or FC for all the bilateral relationships for that institution or country. If institution A has relationships with 2 others, B and C, then the collaboration score is the sum of FC for A + B and A + C. ■

## NATUREINDEX.COM

*A global indicator of high-quality research*



# NATURE INDEX TABLES

Tables are calculated based on collaboration score, which is the sum of the fractional count (FC) for collaborative papers involving the named country and each of its partner countries.

## TOP 100 COUNTRIES BY COLLABORATION SCORE

2014	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
1	United States	29,442	37,042
2	Germany	13,348	17,690
3	United Kingdom	12,111	16,258
4	France	8,289	11,318
5	China	8,830	9,713
6	Italy	4,588	7,188
7	Spain	4,333	6,476
8	Japan	5,241	6,292
9	Canada	4,870	6,105
10	Australia	4,009	5,561
11	Switzerland	4,619	5,390
12	Netherlands	3,606	5,064
13	Sweden	2,695	3,299
14	South Korea	2,445	2,845
15	Belgium	1,942	2,657
16	Denmark	1,874	2,534
17	Russia	1,911	2,428
18	Chile	771	2,237
19	Austria	1,610	1,997
20	Israel	1,465	1,781
21	India	1,343	1,745
22	Brazil	1,054	1,517
23	Singapore	1,506	1,506
24	Poland	992	1,446
25	Finland	1,106	1,431
26	Taiwan	1,042	1,385
27	Portugal	758	1,020
28	Norway	936	993
29	South Africa	557	962
30	Saudi Arabia	805	887
31	Mexico	484	858
32	Czech Republic	601	773
33	Ireland	606	736
34	Greece	480	688
35	Argentina	499	682
36	New Zealand	560	661
37	Hungary	510	631
38	Turkey	353	446
39	Ukraine	277	407
40	Estonia	229	267
41	Slovenia	197	241
42	Croatia	177	240
43	Thailand	194	220
44	Iran	173	220
45	Egypt	191	200
46	Romania	148	174
47	Iceland	141	171
48	Bulgaria	101	155
49	Slovakia	109	145
50	Colombia	126	142

2014	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
51	Lithuania	84	131
52	Malaysia	122	124
53	United Arab Emirates	99	114
54	Panama	107	107
55	Serbia	75	103
56	Peru	90	90
57	Qatar	60	81
58	Venezuela	50	81
59	Luxembourg	79	79
60	Indonesia	78	78
61	Pakistan	69	78
62	Kenya	77	77
63	Vietnam	62	74
64	Armenia	51	70
65	Costa Rica	67	68
66	Cyprus	59	63
67	Ecuador	57	58
68	Georgia	23	56
69	Uruguay	47	49
70	Morocco	43	46
71	Belarus	45	45
72	Algeria	34	44
73	Vatican City State (Holy See)	10	40
74	Philippines	39	39
75	Bangladesh	34	34
76	Lebanon	30	34
77	Uganda	33	33
78	Bolivia	32	32
79	Latvia	28	32
80	Tunisia	25	29
81	Iraq	23	28
82	Cameroon	27	27
83	Uzbekistan	16	27
84	Moldova	26	26
85	Cuba	23	25
86	Nepal	23	23
87	Sri Lanka	21	23
88	Tanzania	21	21
89	Congo	20	20
90	Jordan	20	20
91	Kazakhstan	17	20
92	Oman	17	19
93	Greenland	19	19
94	Nigeria	16	19
95	Kuwait	16	18
96	Benin	16	16
97	Macedonia	15	16
98	Palestine	16	16
99	Mongolia	15	15
100	Ethiopia	15	15

For more information on collaboration score and other metrics, see page S83.

## TOP 100 INSTITUTIONS BY COLLABORATION SCORE

2014	INSTITUTION	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
1	French National Centre for Scientific Research (CNRS)	France	6,458	8,367
2	Max Planck Society	Germany	3,168	5,054
3	Chinese Academy of Sciences (CAS)	China	3,583	4,060
4	Harvard University	United States	3,124	3,975
5	Spanish National Research Council (CSIC)	Spain	1,946	2,950
6	Helmholtz Association of German Research Centres	Germany	2,569	2,684
7	National Institute for Astrophysics (INAF)	Italy	495	2,263
8	University of Cambridge	United Kingdom	1,564	2,126
9	Massachusetts Institute of Technology (MIT)	United States	1,823	2,054
10	Pierre and Marie Curie University (UPMC) - Paris 6	France	1,371	1,953
11	California Institute of Technology (Caltech)	United States	976	1,928
12	University of California Berkeley (UC Berkeley)	United States	1,372	1,848
13	University of Oxford	United Kingdom	1,505	1,817
14	National Aeronautics and Space Administration (NASA)	United States	759	1,792
15	Stanford University (SU)	United States	1,457	1,767
16	The University of Tokyo (UTokyo)	Japan	1,376	1,716
17	Atomic Energy and Alternative Energies Commission (CEA)	France	1,284	1,684
18	Russian Academy of Sciences (RAS)	Russia	1,360	1,624
19	Paris Diderot University (Paris 7)	France	974	1,552
20	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	1,229	1,414
21	University College London (UCL)	United Kingdom	1,139	1,408
22	Columbia University in the City of New York (CU)	United States	1,053	1,269
23	Peking University (PKU)	China	1,130	1,265
24	Yale University	United States	1,015	1,264
25	The University of Edinburgh	United Kingdom	912	1,237
26	Imperial College London (ICL)	United Kingdom	1,133	1,220
27	Lawrence Berkeley National Laboratory (LBNL)	United States	1,071	1,216
28	Smithsonian Institution	United States	390	1,211
29	University of Toronto (U of T)	Canada	986	1,182
30	University of Washington (UW)	United States	983	1,150
31	National Institutes of Health (NIH)	United States	1,143	1,143
32	University of California, San Diego (UC San Diego)	United States	1,054	1,134
33	University of Maryland, College Park (UMCP)	United States	837	1,123
34	The Johns Hopkins University (JHU)	United States	816	1,110
35	University of Michigan (U-M)	United States	818	1,104
36	The University of Chicago (UChicago)	United States	817	1,088
37	Heidelberg University (Uni Heidelberg)	Germany	770	1,068
38	University of Copenhagen (UCPH)	Denmark	798	1,060
39	European Southern Observatory (ESO)	Germany	229	1,042
40	Joseph Fourier University (UJF)	France	792	1,007
41	National Institute for Health and Medical Research (INSERM)	France	1,006	1,006
42	Institute of Physical and Chemical Research (RIKEN)	Japan	963	1,001
43	Leibniz Association	Germany	743	995
44	National Research Council (CNR)	Italy	947	972
45	University of California Los Angeles (UCLA)	United States	841	972
46	Swiss Federal Institute of Technology in Lausanne (EPFL)	Switzerland	918	960
47	Ludwig Maximilian University of Munich (LMU)	Germany	772	958
48	University of Barcelona (UB)	Spain	727	939
49	Cornell University	United States	806	939
50	Kyoto University	Japan	774	909



## TOP 100 INSTITUTIONS BY COLLABORATION SCORE

2014	INSTITUTION	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
51	University of Paris Sud (Paris 11)	France	788	903
52	École Normale Supérieure (ENS)	France	676	887
53	The University of Manchester (UoM)	United Kingdom	696	879
54	University of California Santa Cruz (UCSC)	United States	488	878
55	The University of Arizona (Arizona)	United States	471	876
56	Princeton University	United States	643	873
57	University of Wisconsin-Madison (UW-Madison)	United States	739	868
58	The Pennsylvania State University (Penn State)	United States	524	859
59	Aix-Marseille University (AMU)	France	565	857
60	The Ohio State University (OSU)	United States	620	854
61	University of Glasgow	United Kingdom	679	836
62	University of Padova (UNIPD)	Italy	660	829
63	The University of Sydney (USYD)	Australia	529	825
64	University of Amsterdam (UvA)	Netherlands	603	816
65	University of Bologna (UNIBO)	Italy	639	810
66	Autonomous University of Barcelona (UAB)	Spain	609	801
67	Sapienza University of Rome	Italy	730	798
68	Northwestern University (NU)	United States	730	792
69	Australian National University (ANU)	Australia	424	784
70	Leiden University	Netherlands	369	779
71	University of Colorado Boulder (CU-Boulder)	United States	583	777
72	University of Bonn (Uni Bonn)	Germany	538	776
73	Tsinghua University (TH)	China	752	771
74	The University of Melbourne (UniMelb)	Australia	650	758
75	Osaka University	Japan	710	752
76	Technical University Munich (TUM)	Germany	686	751
77	University of Pennsylvania (Penn)	United States	686	740
78	University of Geneva (UNIGE)	Switzerland	516	739
79	Institute of Research for Development (IRD)	France	578	735
80	Science and Technology Facilities Council (STFC)	United Kingdom	537	735
81	University of Illinois at Urbana-Champaign (UIUC)	United States	647	728
82	Duke University	United States	724	727
83	Academia Sinica	Taiwan	483	726
84	Argonne National Laboratory (ANL)	United States	676	723
85	The University of Texas at Austin (UT Austin)	United States	508	721
86	Paris Observatory	France	195	720
87	University of California Santa Barbara (UCSB)	United States	589	720
88	The University of British Columbia (UBC)	Canada	589	704
89	The University of Warwick (Warwick)	United Kingdom	505	697
90	Utrecht University (UU)	Netherlands	662	692
91	University of Science and Technology of China (USTC)	China	630	690
92	University of Toulouse	France	510	687
93	University of Groningen (RUG)	Netherlands	466	687
94	McGill University	Canada	571	678
95	University of Minnesota (UMN)	United States	574	675
96	Space Telescope Science Institute (STScI)	United States	147	668
97	Autonomous University of Madrid (UAM)	Spain	525	665
98	Polish Academy of Sciences (PAS)	Poland	496	665
99	Radboud University Nijmegen (RU)	Netherlands	530	659
100	University of Hawai'i at Manoa (UH Mānoa)	United States	331	658

## TOP 50 INSTITUTIONS BY COLLABORATION SCORE IN CHEMISTRY

2014	INSTITUTION	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
1	Chinese Academy of Sciences (CAS)	China	1,514	1,514
2	French National Centre for Scientific Research (CNRS)	France	1,370	1,370
3	Max Planck Society	Germany	545	545
4	Peking University (PKU)	China	422	422
5	Spanish National Research Council (CSIC)	Spain	373	373
6	Helmholtz Association of German Research Centres	Germany	370	370
7	University of California Berkeley (UC Berkeley)	United States	368	368
8	National Research Council (CNR)	Italy	366	366
9	Lawrence Berkeley National Laboratory (LBNL)	United States	332	332
10	University of Chinese Academy of Sciences (UCAS)	China	298	298
11	Northwestern University (NU)	United States	296	296
12	Harvard University	United States	265	265
13	Stanford University (SU)	United States	242	242
14	University of Science and Technology of China (USTC)	China	235	235
15	Massachusetts Institute of Technology (MIT)	United States	234	234
16	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	233	233
17	University of Oxford	United Kingdom	230	230
18	Tsinghua University (TH)	China	219	219
19	Swiss Federal Institute of Technology in Lausanne (EPFL)	Switzerland	216	216
20	Argonne National Laboratory (ANL)	United States	213	213
21	The University of Tokyo (UTokyo)	Japan	211	211
22	Nanyang Technological University (NTU)	Singapore	208	208
23	University of Cambridge	United Kingdom	203	203
24	Kyoto University	Japan	200	200
25	Nanjing University (NJU)	China	194	194
26	National University of Singapore (NUS)	Singapore	190	190
27	Atomic Energy and Alternative Energies Commission (CEA)	France	189	189
28	Joseph Fourier University (UJF)	France	178	178
29	Institute of Chemical Research of Catalonia (ICIQ)	Spain	177	177
30	University of Illinois at Urbana-Champaign (UIUC)	United States	171	171
31	National Institute of Applied Sciences (INSA)	France	170	170
32	Nankai University (NKU)	China	169	169
33	University of Strasbourg	France	167	167
34	Pierre and Marie Curie University (UPMC) - Paris 6	France	163	163
35	Leibniz Association	Germany	163	163
36	Tohoku University	Japan	161	161
37	Fudan University	China	160	160
38	University of Milan (UNIMI)	Italy	156	156
39	Osaka University	Japan	152	152
40	Technical University Munich (TUM)	Germany	151	151
41	Russian Academy of Sciences (RAS)	Russia	148	148
42	Xiamen University (XMU)	China	147	147
43	University of Barcelona (UB)	Spain	143	143
44	University of Toulouse	France	141	141
45	Tianjin University (TJU)	China	141	141
46	Korea Advanced Institute of Science and Technology (KAIST)	South Korea	141	141
47	The University of Manchester (UoM)	United Kingdom	139	139
48	Zhejiang University (ZJU)	China	139	139
49	Ludwig Maximilian University of Munich (LMU)	Germany	134	134
50	Agency for Science, Technology and Research (A*STAR)	Singapore	133	133

## TOP 50 INSTITUTIONS BY COLLABORATION SCORE IN EARTH AND ENVIRONMENTAL SCIENCES

2014	INSTITUTION	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
1	French National Centre for Scientific Research (CNRS)	France	804	804
2	Institute of Research for Development (IRD)	France	400	400
3	National Aeronautics and Space Administration (NASA)	United States	363	363
4	Helmholtz Association of German Research Centres	Germany	300	300
5	Chinese Academy of Sciences (CAS)	China	293	293
6	Paris Diderot University (Paris 7)	France	270	270
7	National Oceanic and Atmospheric Administration (NOAA)	United States	259	259
8	California Institute of Technology (Caltech)	United States	248	248
9	Pierre and Marie Curie University (UPMC) - Paris 6	France	213	213
10	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	190	190
11	Centre National d'Etudes Spatiales (CNES)	France	190	190
12	École Normale Supérieure (ENS)	France	188	188
13	University of Colorado Boulder (CU-Boulder)	United States	183	183
14	University of California, San Diego (UC San Diego)	United States	176	176
15	Columbia University in the City of New York (CU)	United States	169	169
16	University of Maryland, College Park (UMCP)	United States	154	154
17	Spanish National Research Council (CSIC)	Spain	153	153
18	Woods Hole Oceanographic Institution (WHOI)	United States	150	150
19	Max Planck Society	Germany	150	150
20	Atomic Energy and Alternative Energies Commission (CEA)	France	144	144
21	University of Washington (UW)	United States	137	137
22	University of California Berkeley (UC Berkeley)	United States	135	135
23	U.S. Geological Survey (USGS)	United States	135	135
24	Paris-Est Créteil University (UPEC)	France	134	134
25	Massachusetts Institute of Technology (MIT)	United States	131	131
26	National Center for Atmospheric Research (NCAR)	United States	124	124
27	École Polytechnique	France	122	122
28	University of Versailles Saint-Quentin-en-Yvelines (UVSQ)	France	121	121
29	University of Cambridge	United Kingdom	118	118
30	Leibniz Association	Germany	117	117
31	University of Oxford	United Kingdom	116	116
32	The University of Tokyo (UTokyo)	Japan	114	114
33	The Commonwealth Scientific & Industrial Research Organisation (CSIRO)	Australia	110	110
34	Japan Agency for Marine-Earth Science and Technology (JAMSTEC)	Japan	106	106
35	University of Hawai'i at Manoa (UH Mānoa)	United States	105	105
36	Joseph Fourier University (UJF)	France	105	105
37	The University of Texas at Austin (UT Austin)	United States	104	104
38	Princeton University	United States	97	97
39	Peking University (PKU)	China	96	96
40	University of Wisconsin-Madison (UW-Madison)	United States	95	95
41	Australian National University (ANU)	Australia	94	94
42	University of Tasmania (UTAS)	Australia	93	93
43	The University of Arizona (Arizona)	United States	89	89
44	Harvard University	United States	89	89
45	University of Toulouse	France	89	89
46	French Research Institute for Exploitation of the Sea (Ifremer)	France	86	86
47	University of Minnesota (UMN)	United States	86	86
48	University of Bristol (UoB)	United Kingdom	84	84
49	University of Oslo (UiO)	Norway	83	83
50	The Pennsylvania State University (Penn State)	United States	83	83

## TOP 50 INSTITUTIONS BY COLLABORATION SCORE IN LIFE SCIENCES

2014	INSTITUTION	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
1	Harvard University	United States	2,192	2,192
2	French National Centre for Scientific Research (CNRS)	France	1,523	1,523
3	National Institutes of Health (NIH)	United States	1,089	1,089
4	Max Planck Society	Germany	1,051	1,051
5	National Institute for Health and Medical Research (INSERM)	France	899	899
6	Massachusetts Institute of Technology (MIT)	United States	787	787
7	Helmholtz Association of German Research Centres	Germany	762	762
8	University of Oxford	United Kingdom	662	662
9	Chinese Academy of Sciences (CAS)	China	661	661
10	University College London (UCL)	United Kingdom	641	641
11	University of Cambridge	United Kingdom	628	628
12	Stanford University (SU)	United States	592	592
13	University of California San Francisco (UCSF)	United States	573	573
14	University of Toronto (U of T)	Canada	562	562
15	University of California, San Diego (UC San Diego)	United States	544	544
16	Institute of Physical and Chemical Research (RIKEN)	Japan	542	542
17	Columbia University in the City of New York (CU)	United States	517	517
18	Yale University	United States	515	515
19	University of Washington (UW)	United States	513	513
20	Imperial College London (ICL)	United Kingdom	500	500
21	Cornell University	United States	470	470
22	Duke University	United States	466	466
23	University of Copenhagen (UCPH)	Denmark	424	424
24	The Johns Hopkins University (JHU)	United States	418	418
25	University of Pennsylvania (Penn)	United States	411	411
26	Spanish National Research Council (CSIC)	Spain	410	410
27	University of California Berkeley (UC Berkeley)	United States	402	402
28	University of California Los Angeles (UCLA)	United States	399	399
29	King's College London (KCL)	United Kingdom	394	394
30	The University of Tokyo (UTokyo)	Japan	387	387
31	Washington University in St. Louis (WUSTL)	United States	387	387
32	Karolinska Institute (KI)	Sweden	381	381
33	University of Michigan (U-M)	United States	364	364
34	Pierre and Marie Curie University (UPMC) - Paris 6	France	358	358
35	Baylor College of Medicine (BCM)	United States	350	350
36	Ludwig Maximilian University of Munich (LMU)	Germany	341	341
37	The University of Chicago (UChicago)	United States	336	336
38	Icahn School of Medicine at Mount Sinai (ISMMS)	United States	325	325
39	Wellcome Trust Sanger Institute	United Kingdom	322	322
40	French National Institute for Agricultural Research (INRA)	France	321	321
41	University of North Carolina at Chapel Hill (UNC)	United States	317	317
42	Istituto Di Ricovero e Cura a Carattere Scientifico (IRCCS)	Italy	307	307
43	The University of Texas MD Anderson Cancer Center	United States	302	302
44	Technical University Munich (TUM)	Germany	300	300
45	Vanderbilt University (VU)	United States	296	296
46	Utrecht University (UU)	Netherlands	295	295
47	New York University (NYU)	United States	294	294
48	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	294	294
49	The University of Edinburgh	United Kingdom	293	293
50	The University of Melbourne (UniMelb)	Australia	292	292



## TOP 50 INSTITUTIONS BY COLLABORATION SCORE IN PHYSICAL SCIENCES

2014	INSTITUTION	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
1	French National Centre for Scientific Research (CNRS)	France	3,194	5,103
2	Max Planck Society	Germany	1,731	3,616
3	National Institute for Astrophysics (INAF)	Italy	491	2,260
4	Spanish National Research Council (CSIC)	Spain	1,169	2,173
5	Chinese Academy of Sciences (CAS)	China	1,468	1,946
6	Harvard University	United States	828	1,678
7	California Institute of Technology (Caltech)	United States	557	1,508
8	National Aeronautics and Space Administration (NASA)	United States	394	1,427
9	Helmholtz Association of German Research Centres	Germany	1,302	1,417
10	Russian Academy of Sciences (RAS)	Russia	1,091	1,355
11	University of Cambridge	United Kingdom	760	1,323
12	Pierre and Marie Curie University (UPMC) - Paris 6	France	722	1,305
13	Atomic Energy and Alternative Energies Commission (CEA)	France	821	1,221
14	University of California Berkeley (UC Berkeley)	United States	713	1,189
15	Massachusetts Institute of Technology (MIT)	United States	872	1,103
16	Paris Diderot University (Paris 7)	France	513	1,091
17	The University of Tokyo (UTokyo)	Japan	741	1,081
18	Smithsonian Institution	United States	258	1,079
19	European Southern Observatory (ESO)	Germany	229	1,042
20	Stanford University (SU)	United States	729	1,039
21	University of Oxford	United Kingdom	640	951
22	University of Maryland, College Park (UMCP)	United States	557	843
23	Lawrence Berkeley National Laboratory (LBNL)	United States	688	833
24	The University of Edinburgh	United Kingdom	492	818
25	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	620	805
26	Heidelberg University (Uni Heidelberg)	Germany	451	749
27	Paris Observatory	France	181	706
28	The University of Arizona (Arizona)	United States	285	689
29	Joseph Fourier University (UJF)	France	465	679
30	University of Padova (UNIPD)	Italy	505	673
31	Science and Technology Facilities Council (STFC)	United Kingdom	476	673
32	Sapienza University of Rome	Italy	600	668
33	Space Telescope Science Institute (STScI)	United States	147	668
34	University of California Santa Cruz (UCSC)	United States	277	666
35	University of Paris Sud (Paris 11)	France	544	659
36	University of Michigan (U-M)	United States	373	659
37	University College London (UCL)	United Kingdom	381	651
38	University of Bologna (UNIBO)	Italy	479	649
39	Princeton University	United States	418	648
40	Peking University (PKU)	China	509	643
41	The University of Chicago (UChicago)	United States	364	634
42	University of Barcelona (UB)	Spain	416	628
43	The Ohio State University (OSU)	United States	394	627
44	Yale University	United States	377	625
45	University of Glasgow	United Kingdom	468	625
46	Swiss Federal Institute of Technology in Lausanne (EPFL)	Switzerland	578	620
47	The University of Sydney (USYD)	Australia	317	613
48	The Johns Hopkins University (JHU)	United States	312	606
49	Imperial College London (ICL)	United Kingdom	516	603
50	Aix-Marseille University (AMU)	France	307	599

## TOP 50 INDUSTRIAL ORGANIZATIONS BY COLLABORATION SCORE

2014	CORPORATION	COUNTRY	WEIGHTED COLLABORATION SCORE	COLLABORATION SCORE
1	BGI	China	137	137
2	IBM Corporation	United States	80	90
3	F. Hoffman-La Roche AG	Switzerland	85	85
4	Samsung Group	South Korea	75	75
5	Novartis International AG	Switzerland	66	66
6	GlaxoSmithKline plc	United Kingdom	63	63
7	AstraZeneca plc	United Kingdom	63	63
8	Pfizer Inc.	United States	57	57
9	Interuniversity Microelectronics Centre (IMEC)	Belgium	55	55
10	Amgen Inc.	United States	51	51
11	Merck & Co., Inc.	United States	37	37
12	GNS Science	New Zealand	36	36
13	Microsoft Corporation	United States	31	33
14	Toyota Group	Japan	30	30
15	Leidos Holdings, Inc.	United States	30	30
16	Lockheed Martin Corporation	United States	14	29
17	Polyera Corporation	United States	28	28
18	Eureka Scientific, Inc.	United States	6	28
19	European XFEL GmbH	Germany	25	26
20	Adler Planetarium	United States	5	26
21	Boehringer Ingelheim GmbH	Germany	25	25
22	Thales S.A.	France	25	25
23	Bruker Corporation	United States	25	25
24	Raytheon Company	United States	6	24
25	Thermo Fisher Scientific Inc.	United States	24	24
26	BASF SE	Germany	23	23
27	Hitachi, Ltd.	Japan	22	22
28	NTT Group	Japan	21	21
29	Royal Philips (Philips)	Netherlands	21	21
30	Agilent Technologies, Inc.	United States	20	20
31	Sanofi	France	19	19
32	Eli Lilly and Company	United States	19	19
33	Illumina, Inc.	United States	17	17
34	Takeda Pharmaceutical Company Limited	Japan	17	17
35	Wuhan Research Institute of Posts and Telecommunications (WRI)	China	17	17
36	Mitsubishi Group	Japan	16	16
37	Cell Signaling Technology, Inc.	United States	16	16
38	Haldor Topsøe A/S	Denmark	16	16
39	NYCB Real-Time Computing Inc.	United States	8	15
40	The Dow Chemical Company	United States	14	14
41	Aerodyne Research, Inc. (ARI)	United States	14	14
42	Eni s.p.a.	Italy	14	14
43	FEI Company	United States	14	14
44	Johnson & Johnson	United States	14	14
45	SRI International	United States	12	13
46	Western Digital Corporation	United States	13	13
47	Novo Nordisk A/S	Denmark	13	13
48	Predictive Science Inc.	United States	2	12
49	Element Six	Luxembourg	12	12
50	China Electronics Technology Group Corporation (CETC)	China	12	12

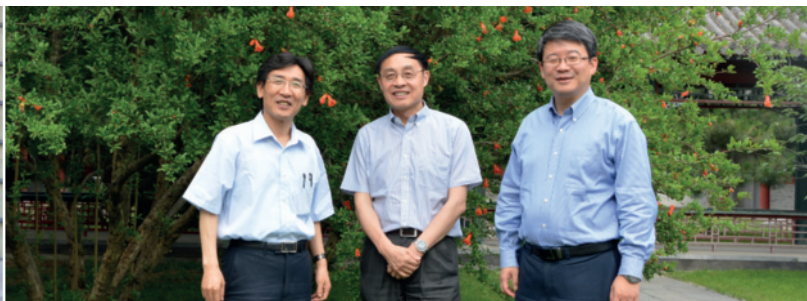
# Powering population studies. Advancing healthcare.



Improving healthcare is a global challenge. As a truly global company, BGI is in a unique position to provide solutions to meet that challenge. Solutions like our comprehensive range of sequencing services, extensive bioinformatics expertise, and new high-capacity Revolocuity™ system, a whole genome, whole exome sequencing platform that enable researchers to explore human genetic variation at a scale never before possible. And make discoveries that advance healthcare for the benefit of us all.

Learn more at [www.bgi-international.com](http://www.bgi-international.com)





Collaborative Innovation Center of Quantum Matter:

# Highlighting the quantum golden triangle

**The Collaborative Innovation Center of Quantum Matter (CICQM) was founded on the premise that to tackle ambitious challenges in quantum matter research, it is vital to bring together world-class scientists and exceptional technological resources from different organizations.**

CICQM was established in 2012 as a joint venture of Peking University, Tsinghua University and the Institute of Physics at the Chinese Academy of Sciences, in line with the mission and vision of the National 2011 Plan. Initiated by the Ministry of Education and Ministry of Finance of China, the 'National 2011 Plan' is a topically diverse, centre-based programme supporting multi-institutional collaborations in research and education of the highest quality. CICQM was launched in 2013 as one of the 14 proposals selected in the first round of the National 2011 Plan.

CICQM has its roots in decades of fruitful collaborations among physicists at Peking University, Tsinghua University and the Institute of Physics. These three partner institutions each have excellent individual reputations in the physical sciences and are located in the Zhongguancun region of Beijing, about two miles away from each other. As a result of these advantages, CICQM is becoming the core of this fast-growing 'quantum golden triangle', combining the specific strengths of the partner institutions in a new organization

that has an open and deeply collaborative atmosphere.

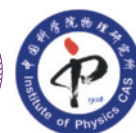
CICQM aims to make new discoveries by focusing on fundamental problems in quantum matter science and to provide leadership for the country pertaining to new quantum materials and phenomena that could ultimately address national needs, including innovation and sustainability. CICQM encompasses six research areas: novel quantum and topological states; high-temperature superconductivity; quantum structures and properties in low-dimensional systems; mesoscopic optical and cold atomic/molecular physics; many-body and strongly correlated quantum systems; and the development of experimental facilities and techniques. In each area, collaborations are highly encouraged between partner institution members and also at an international level.

In the period 2013–2014, CICQM had 431 Science Citation Index publications, many of which appeared in prestigious scientific journals such as *Science*, *Nature* journals and *Physical Review Letters*. In addition, CICQM members have harvested many national and international awards, including the Benjamin Franklin Medal, the Asian Union of Magnetism Societies Award, the Elsevier Most Cited Chinese Researchers Award and the State Natural Science Award.

Led by Xin-Cheng Xie, Qi-Kun Xue and Hong Ding (pictured above right), CICQM has about 170 members from the three partner institutions. Of these team

members, 17 are members of the Chinese Academy of Sciences, 12 are winners of the National Thousand Talent Plan for Global Experts, 52 are winners of the National Science Foundation of China Program for Distinguished Young Scholars, 20 are Cheung Kong Scholars and 34 are winners of the National Thousand Talent Plan for Young Scholars.

CICQM is looking for new researchers of all levels to join our team. Successful applicants will hold primary appointments in the School of Physics at Peking University, the Department of Physics at Tsinghua University or the Institute of Physics at CAS and will have the same rights and responsibilities as other members of these institutions. Positions include distinguished chair professors, tenured full/associate professors, tenure-track associate/assistant professors and postdoctoral fellows. Applicants selected by the CICQM Search Committee based on its own cross-institutional standards will gain supplementary financial support and full access to CICQM's scientific platforms to extend their academic capabilities and opportunities.



## Contact

**Collaborative Innovation Center of Quantum Matter**

**Contact us: [cicqm2011@pku.edu.cn](mailto:cicqm2011@pku.edu.cn)**





# :insideview

## profile feature



**Xiaobing Ren**, *Chairman of Frontier Institute of Science and Technology, Xi'an Jiaotong University*

*Frontier Institute of Science and Technology (FIST) was established in December 2010 with an important investment from Xi'an Jiaotong University (XJTU). Within just five years, FIST has set up 11 research centres of excellence in physics, chemistry, bio-science/life science/basic medical science, and materials science. Xiaobing Ren, a scientist in smart materials and the founding Chairman of FIST, discusses here how FIST was established and shares his plans for further development of the institute.*

**Q: How did you come up with the idea to establish FIST at XJTU?**

I started a collaborative working relationship with XJTU back in 2002-2003, first as a *Cheung Kong* Adjunct Professor giving lectures on campus. Then, I built the Multidisciplinary Materials Research Centre. When I was selected as a member of the national '1,000 Talents Plan' in 2009, I anticipated that we would start to see an immense growth in both the quantity and quality of scientists coming back to China. So I proposed that XJTU should seize this opportunity to attract high-level researchers and build a world-class, multidisciplinary research institute. FIST was one of the first research institutes of its kind in China.

**Q: What are some the major achievements since the foundation of FIST?**

In terms of academic accomplishments, the high-quality faculty team and research are the major achievements. We have also established a novel academic and administrative management system that has turned out to be very effective. In the first two years we focused on faculty recruitment and setting up labs. We now have 11 research centres and over 40 faculty members, most with overseas training experience. The quick increase in talent has brought growth in both the quantity and quality of research output. We published 162 journal articles last year, a 60% increase from the year before. FIST researchers were the first or corresponding authors of approximately 130 of these papers, and nearly 20 were published in top-ranking international journals. In just the first 10 months of 2015, FIST researchers published 28 papers in top-ranking international journals—a 40% increase as compared to the entire year of 2014. As one example of how the research conducted at FIST

contributes to the development of basic science subjects at XJTU, the research led by Min Zhuo, published in *Neuron*, clarified the mechanism of the interactions between anxiety and chronic pain. This research has great potential for clinical applications, and it was the first publication in *Neuron* to list XJTU as a home institution of the corresponding author.

“

A unique management system and an interdisciplinary approach set FIST apart from its domestic counterparts.

”

**Q: What factors do you think contribute most to the success of FIST?**

A unique management system and an interdisciplinary approach set FIST apart from its domestic counterparts. Following international practices, we have adopted the tenure-track system and a transparent faculty evaluation system with more nuanced assessment criteria for tenure promotion. While these may be common in Western institutions, they are innovative in China. FIST is also novel in its wide coverage of subject areas with a faculty size similar to that of a university department. Even within the same research centre, faculty members may have very different research backgrounds. By bringing researchers of different fields together to investigate an issue and exchange ideas, we can broaden their views and gain new perspectives. Our students also benefit from interdisciplinary training by participating in these cross-disciplinary research projects.

**Q: How do you characterize the education program at FIST?**

As just mentioned, we emphasize the interdisciplinary approach in student training to help broaden student views. Graduate students primarily learn by working on research projects—many of which are interdisciplinary—through which they identify study needs and become motivated to study on their own. Another feature

of FIST is its international flavour. As most of our faculty have overseas backgrounds, we can offer seminars in English and ask students to deliver English presentations and reports. Students also have ample opportunities to attend international conferences and short-term training abroad. For instance, one of my students delivered an invited talk at the Energy Materials Nanotechnology Spring Meeting held in the United States in 2013. Some of our students have also published in high-quality international journals as corresponding authors, demonstrating their high research capabilities.

**Q: How do you find a balance between conducting research and managing the whole institute?**

The institute is not managed solely by me. I am able to continue conducting research thanks to our capable administrative staff. I am primarily in charge of strategic decisions and planning, while Vice Chairman Dr. Xiangli Meng is responsible for the daily management of the institute. The vice chairman position is a full-time administrative role requiring rich management experience as well as overseas research experience—it is important for our administrators to understand science and research so that they can easily communicate with our faculty members and researchers to address burning issues. We also emphasize transparency and democracy in our policy planning, and all faculty members are encouraged to participate in policy making for the institute. In this way, managing the institute is a group effort.

**Q: What research fields will FIST focus on in its next stage of development?**

We do not set particular focus areas at FIST, as we encourage a multidisciplinary approach. Hot research fields today may not be trendy tomorrow. By conducting multidisciplinary research, we can get innovative ideas and identify new trends and new hotspots. We want to further broaden our research fields by attracting more researchers in math and computer science, biomedical and life sciences, and physics and material sciences. Our five-year goal is to triple our faculty size and become an internationally influential institute of excellence in interdisciplinary research.



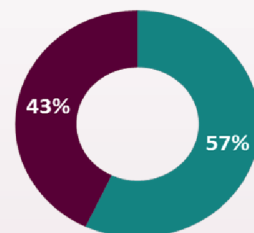
Frontier Institute of Science and Technology (FIST) is an important investment by Xi'an Jiaotong University (XJTU) in an effort to establish a world-class, multi-disciplinary research institute.

FIST has built up a multi-discipline instrument sharing platform valued more than one hundred million RMB yuan. The research fields in FIST cover physics, chemistry, biology (including life sciences and basic medicine), materials science, mathematics, computation, and basic engineering, etc.

After nearly five-year development, FIST has obtained great achievements, and now it is renowned at home and abroad and has made an important contribution to science disciplines in XJTU.

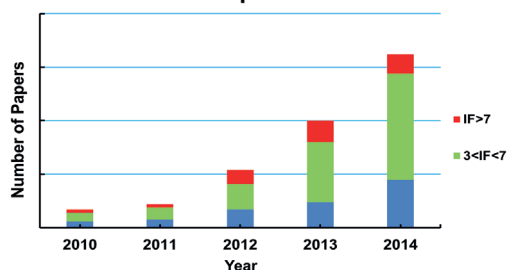
## Faculty

- Since its establishment, FIST has attracted thousands of scholars from all over the world to apply for its positions. After strict selection, 44 scholars have joined FIST, including 4 foreign academicians, 9 "1000 Talent Plan" scholars, 6 "1000 Young Talent Plan" scholars, and 1 scholar of NSFC excellent young scientist foundation.
- In FIST, 43% of faculty members are academicians or have national titles. The talent quality and density have reached the top level of the colleges in China.



■ Academicians & faculty members who were awarded national titles

FIST Paper statistics



## Publications

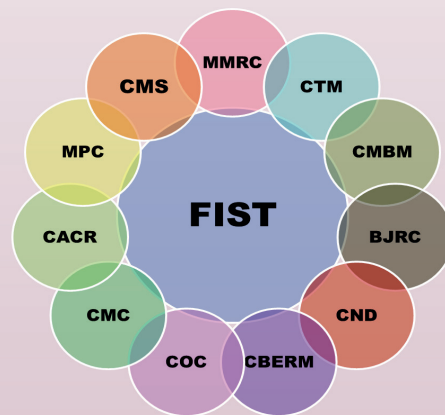
After nearly five-year development, FIST has published a number of outstanding papers in scientific research.

- Among the outstanding papers published in 2014, 72% were published in top-level journals with an IF greater than 3.

IF: Impact Factor

## Multi-discipline

- Multi-disciplinary Materials Research Center (MMRC)
- Center of Microstructure Science (CMS)
- Materials Physics Center (MPC)
- Center for Biomedical Engineering and Regenerative Medicine (CBERM)
- Center for Applied Chemical Research (CACR)
- Center for Neuron and Disease (CND)
- Center for Materials Chemistry (CMC)
- Center for Organic Chemistry (COC)
- Center for Mitochondrial Biology and Medicine (CMBM)
- Center for Translational Medicine (CTM)
- Bone and Joints Research Center (BJRC)



## Honors & Awards

- Prof. Pengfei Li in Center for Organic Chemistry was awarded the 2015 Thieme Chemistry Journal Award.
- Prof. Min Zhuo and Prof. Jiankang Liu were selected on Elsevier 2014 Most Cited Chinese Researchers list.
- Prof. Ju Li and Prof. Yadong Yin were selected on Thomson Reuters 2014 Highly Cited Researchers list.
- Prof. Xiaobing Ren's study on Strain Glass has become a hotspot of international research. The TMS2015 set up a special session for this topic.
- The paper published on *Science* by Prof. Ju Li, Prof. Xiaobing Ren and collaborators from China University of Petroleum won the 2013 Ten Major Advances in Science in China.
- FIST was awarded National Advanced Group of Professional and Technical Personnel in 2014.



Gunma University – Heavy-Ion Radiotherapy

# Exploring new frontiers in heavy-ion radiotherapy

**A pioneer of carbon-beam radiotherapy, Gunma University continues to lead the clinical application of heavy-ion beams through the highly successful Gunma University Heavy Ion Medical Research Center and the new Gunma University Initiative for Advanced Research.**

Heavy-ion radiotherapy is a next-generation medical technology that in just a few years has radically transformed the face of cancer treatment. This technique can target and destroy inoperable tumours with three-dimensional submillimetre accuracy in just a few short sessions, without radiation damage to surrounding healthy tissue. For brain cancers and other complicated tumours, heavy-ion therapy is now an effective medical option that promises cancer patients a vastly improved quality of life.

Gunma University, in collaboration with the National Institute of Radiological Sciences, was one of the first to adapt a heavy-ion particle accelerator for use in a clinical environment, and since it commenced operation in 2010, the Gunma University Heavy Ion Medical Research Center (GHMC) has remarkably increased the survival rates of patients with inoperable or complicated cancers. The facility is still one of only two in the world attached to a university hospital, which greatly enhances the cross-disciplinary academic and clinical research and development of carbon beam therapy.

Having demonstrated the medical

efficacy of heavy-ion therapy for cancer treatment in a clinical setting, the university is now turning its sights to further advancing the technology and expanding its range of applications, including microsurgery and even biomedical imaging.

“Last year, we established the Gunma University Initiative for Advanced Research to progress the heavy-ion beam technology further,” says Takashi Nakano, previously centre director and now professor and chairman of the Department of Radiology and Radiation Oncology. “The first goal of the initiative is to develop what we call ‘integrated oncology research’, which combines our core expertise in heavy-ion therapy with our strengths in endocrinology, metabolism and signalling research.”

Nakano’s team is also undertaking fundamental research on the biological basis for the high clinical antitumour effectiveness of carbon-beam therapy. “Using the latest super-resolution microscopy techniques, we have identified the mechanism of high cancer cell lethality, which will help us fine-tune various parameters of the technique,” Nakano explains. “We have also been investigating the relationship between carbon ion sensitivity and cancer-related gene abnormalities, which we also expect to feed back into improved therapies.”

By introducing a rapid dose-shaping fabrication system and modifying the ion beam extraction method to a more advanced radiofrequency knockout configuration, GHMC researchers have significantly shortened the waiting time for treatment,



Professor Takashi Nakano, chairman of the Department of Radiology and Radiation Oncology, School of Medicine, Gunma University.

sharpened the ion energy distribution and cut the treatment time itself. The ultra-sharp three-dimensional dose shapes now achievable make the carbon beam not only more effective for precision tumour treatment in difficult cases, but also suitable for use as a precise tool for microsurgery. “We now have a microsurgery-ready beamline, which will open many new opportunities for treatment,” says Nakano.



## Contact

**Gunma University Graduate School of Medicine**  
**3-39-22 Showa-machi, Maebashi City, Gunma, Japan**  
**Phone: +81 27 220 8383**  
**<http://heavy-ion.showa.gunma-u.ac.jp/en/index.php>**





## Ready to see a glimpse of tomorrow?

Visit KAUST Discovery to see the latest international research aiming to address some of the most pressing challenges facing societies everywhere:

- How can crops grow in sea water?
- How can paper be used as an early warning signal against natural disasters?
- How could enzymes help stop the spread of malaria?

KAUST Discovery highlights the cutting-edge research, technologies and innovations emerging from the King Abdullah University of Science and Technology (KAUST).

From biotechnology, to solar, to materials and marine science: KAUST is working on it.

**For a glimpse of tomorrow  
visit KAUST Discovery today.**

[discovery.kaust.edu.sa](http://discovery.kaust.edu.sa)



# EXCELLENCE IN RESEARCH

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN



As one of Europe's leading research universities, Ludwig-Maximilians-Universität (LMU) in Munich is committed to the highest standards of excellence in research and teaching. Building on a 500-year-tradition of scholarship, LMU covers a broad spectrum of disciplines, ranging from the Humanities and Cultural Studies through Law, Economics and Social Studies to Medicine and the Natural Sciences. The skills and experience of LMU's faculty and staff form the basis for its distinguished record in research, and for the University's consistently high rating in international rankings.

## LMU offers a first-rate academic environment

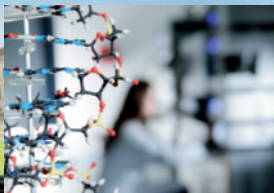
- A unique and diverse academic culture encompassing all areas of learning and research
- State-of-the-art facilities and conditions for innovative research in all fields of knowledge
- An integral part of the Munich research area – Germany's most vibrant location for science and technology

## LMU supports researchers at all career stages

- A structured Academic Career Program for doctoral students and post-docs, enabling them to develop their skills and pursue their research goals in an intellectually stimulating setting
- Attractive positions and a well-established tenure-track model
- An Equal Opportunities program fostering the talents of female researchers and students

## LMU promotes collaboration between scholars from all over the world

- Large-scale collaborative projects in all subject areas – at national and international levels
- Targeted support schemes to facilitate academic exchange, such as LMU's Center for Advanced Studies
- A member of distinguished academic networks of international renown







# :insideview

## profile feature

**Oliver Bogler, Ph.D., Senior Vice President Academic Affairs,**  
The University of Texas MD Anderson Cancer Center

ADVERTISEMENT FEATURE  
THE UNIVERSITY OF TEXAS  
**MD Anderson**  
**Cancer Center**  
Making Cancer History™

MD Anderson Cancer Center was formed in 1941 as part of The University of Texas System and is one of 45 U.S. comprehensive cancer centres as designated by the National Cancer Institute. The institution employs more than 20,000 people and provided care for 127,000 cancer patients in 2014, including more than 8,000 in clinical trials, making it the largest clinical trial programme in the country.

In 2002, MD Anderson launched its Global Academic Programs (GAP), whose mission is to create a global collaborative network dedicated to research, prevention and education. Nature discussed GAP with Oliver Bogler, Ph.D., MD Anderson's Senior Vice President of Academic Affairs. A biologist and brain cancer researcher by trade, Dr. Bogler took the reins at GAP in 2010.

### **Q: How does MD Anderson collaborate with international institutions?**

Our strategy is to follow the faculty, as faculty involvement is the key to successful relationships. Typically a faculty member will tell GAP about existing collaborations, and where they think other opportunities might exist. We then try to broaden this collaboration to include additional MD Anderson faculty from other disciplines. If we are able to turn it into a multi-disciplinary partnership, then we may form a sister institution relationship. We currently have 33 sister institutions in 24 countries on every continent except Africa and Antarctica, and five of these are actually consortia with multiple partners.

### **Q: What do you mean by multi-disciplinary?**

Unlike many parts of the world, where patients may receive cancer care from a single doctor, MD Anderson practices under a model of multi-disciplinary cancer care. We surround patients with experts from many disciplines including medical and radiation oncology, surgery and others, as appropriate for that patient.

We replicate this multi-disciplinary approach in our academic collaborations, which include medical care, research, prevention and education. We have seeded 82 projects over the past four years, and the scope of the work is tremendously broad. There has been research on cell-signaling and genetics, work that directly benefits patients

“The partnerships are bidirectional, meaning we get as much out of them as we put into them.”

like controlling infections in central line catheters, work on palliative care, integrative medicine and everything in-between.

### **Q: What are some examples of these collaborations?**

Two programmes we are excited about revolve around tobacco control, one in partnership with the Instituto de Cancerologia in Medellin, Colombia, and one with Mexico's Instituto Nacional de Cancerologia. In these programmes, GAP is supporting the work of our colleagues who educate citizens about the dangers of tobacco to prevent addiction and cancer. For example, our faculty are deploying a multimedia educational tool called ASPIRE with a message intended to prevent middle school kids from starting to use tobacco. We are tailoring it for Mexico and Colombia's respective cultures and working with authorities there to help reduce the incidence of lung cancer.

### **Q: Are you attempting to make in-roads in Africa?**

Absolutely, and these partnerships are increasingly important because cancer is on the rise in Africa, especially cervical cancer and AIDS-related cancers. We have been in Africa for the past three years and we have been partnering there with U.S. and international government agencies, and non-profit organizations like Pink Ribbon Red Ribbon and the Union for International Cancer Control, as well as hospitals, with a focus on capacity building.

### **Q: What are some other aspects of GAP?**

We focus a lot on education. We have an annual network conference where our faculty exchange research results and leadership ideas with their counterparts from our sister institutions. We also put on regional conferences, and we do a lot of multi-disciplinary patient care video conferences. For example, we host a bimonthly video conference with our partners from Brazil and Chile on urology. Every centre in the group

presents a case, which are all discussed. This isn't telemedicine, mind you, but experts from multiple disciplines sharing insights and ideas to learn from one another. And that is an important aspect of GAP: the partnerships are bidirectional, meaning we get as much out of them as we put into them.

### **Q: What does it take to become a sister institution?**

It requires a lot of institutional commitment from both sides. One of the most important things is having faculty champions at each institution. Without faculty buy-in it won't go anywhere. So when we get an inquiry from external institutions, we look right away for faculty champions from both sides. We are evaluating about a dozen potential sister institutions at any one time, and I'm personally going to travel to look at two or three potential partners in the next few months. If there is a spark, then we try to make a flame, but we need faculty to make that initial spark.

### **Q: What is the Sister Institution Network Fund (SINF)?**

The SINF is the first financial support system we have been able to offer specifically to MD Anderson researchers who are working with sister institutions. The fund is based on global collaborations, but the funding is local. We don't have the means to provide research funds outside of our institution, so our partners provide the funds for the work that is done by their own faculty. These are two-year projects, and we fund US\$100,000 per project. We select around 20 new projects each year, so there are 40 or so projects ongoing at any time. Since launching SINF in 2010, we've invested about US\$8.2 million and the network has matched us. Our faculty have published some great papers from this work, including in *Nature*, and have used this seed mechanism to get follow-up funding from the National Institutes of Health, foundations and the state of Texas.

# ~~“Cancer,~~ meet our worldwide network”

## MD Anderson | Cancer's #1 Enemy

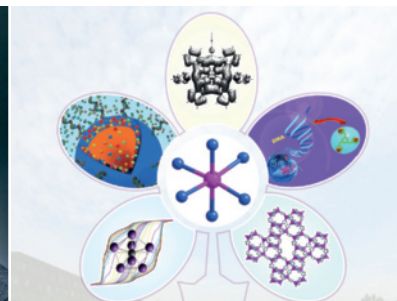
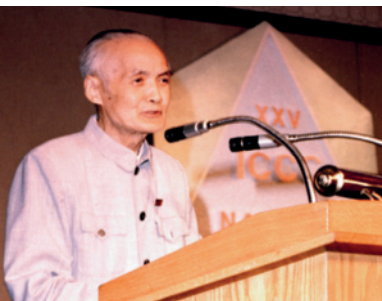
It takes more than a village. To end cancer, it takes a network – a global network of select cancer centers working together on research and education to lessen the world's cancer burden. Because at MD Anderson, we have one goal: to end cancer. For good. To learn more, visit [MakingCancerHistory.com](http://MakingCancerHistory.com).



Ranked number one in the  
nation for cancer care by  
U.S. News & World Report.

THE UNIVERSITY OF TEXAS  
**MD Anderson**  
~~Cancer Center~~

Making Cancer History®



State Key Laboratory of Coordination Chemistry (SKLCC)

# Collaboration boosts coordination chemistry

**The State Key Laboratory of Coordination Chemistry conducts cutting-edge interdisciplinary research at the frontiers of inorganic, organic, polymer and materials chemistry. It is the first state key laboratory for inorganic chemistry in China.**

Emerging from the Institute of Coordination Chemistry, Nanjing University — established in 1978 as the first inorganic chemistry research institute in China — the State Key Laboratory of Coordination Chemistry (SKLCC) was officially approved by the Ministry of Science and Technology in 1991. Led by An-Bang Dai, Xiao-Zeng You and many others, the SKLCC has played a key role in the development of coordination chemistry research in China over the past two and half decades. It achieved an 'A level' laboratory accreditation in the first nation-wide evaluation in 1995 and has been ranked highly in subsequent evaluations.

## Coordination, collaboration and interdisciplinary research

The SKLCC has been focusing on basic coordination chemistry research including synthesis, structures, properties, functions and applications. Its key research areas are based on fundamental science and are closely related to materials and life sciences. These areas include: (1) Functional coordination chemistry; assembly of molecular and supramolecular materials and applications of these materials to information

storage, photoelectric conversion, energy storage, adsorption and separation materials, molecular magnetism, and molecular electronic devices. (2) Bioinorganic chemistry; investigating the synthesis and structure of biologically active complexes and metallic biological macromolecules, designing new metallic medicines, mimicking metal enzymes and so on. (3) Solid-state coordination chemistry; exploring new synthetic methods for preparing low-dimensional, layered, porous nanomaterials. (4) Surface coordination chemistry; studying surface and interface microstructures and their applications in advanced materials. (5) Organometallic chemistry; synthesizing new organometallic compounds based on coordination chemistry, studying new metal complexes in inert bond activation and catalysis, and developing applications for metal organic materials. Using metal organic chemical vapour deposition, the university has realized large-scale industrial production of high-purity metal organic sources and now accounts for more than 80 per cent of the market share in China.

Ever since its foundation, the SKLCC has pursued extensive collaborations and scientific exchanges with domestic and international research teams, resulting in remarkable achievements. The SKLCC has also hosted many national and international conferences, which have had a major impact on the development of coordination chemistry in China. In 1987, the SKLCC hosted the 25th International Conference of Coordination Chemistry — the biggest

international conference held in China since the national economy reforms in the late 1970s. The SKLCC enthusiastically works with scientists around the world to stimulate and promote coordination chemistry research in China.

## Attracting talented researchers

Aspiring to become one of the world's leaders in coordination chemistry research, the SKLCC is very keen to stimulate further cooperation and to intensify collaborations in research projects of mutual interest. Professors, visiting scholars, post-doctoral fellows and graduate students are all encouraged to apply to join the SKLCC. Various funding schemes including generous scholarships, research funds, preferential enrollment policies and other measures have been implemented to facilitate this. A number of outstanding young scholars have been recruited by the SKLCC in the past few years, six of whom were selected in the Thousand Youth Talents Plan in China.



南京大學  
NANJING UNIVERSITY

Visit:  
Email:

<http://sklcc.nju.edu.cn/>  
[sklc@nju.edu.cn](mailto:sklc@nju.edu.cn)





## National Center for Liver Cancer

# A centre for liver cancer research in Shanghai

**The National Center for Liver Cancer set to open in 2015 will foster research and innovation for the early diagnosis, treatment and prevention of liver cancer**

Liver cancer ranks as the second leading cause of cancer-related mortality in developing countries and the sixth in developed countries. An estimated 782,500 new cases of liver cancer and 745,500 deaths occurred worldwide in 2012, of which China alone accounted for about 50 per cent. An extraordinarily heterogeneous disease, liver cancer has become a serious social and medical issue, especially in China.

To address this enormous challenge, China's National Development and Reform Commission established the National Center for Liver Cancer, which is set to open in Shanghai at the end of 2015. The centre aims to strengthen China's capacity in frontier research and clinical treatment of liver cancer and enhance the competitiveness of China's medical industry for treating liver cancer. It will focus on research into biomedicine and translational medicine, as well as fostering

self-innovation in its pursuit of the study of hepatobiliary diseases, especially liver cancer. It is the second national centre in the field of biotechnology established by the commission.

Initial plans for the centre came out of a proposal to promote the scientific study of liver cancer submitted by Mengchao Wu and Hongyang Wang, academicians of the Chinese Academy of Sciences and Chinese Academy of Engineering, respectively, along with five other renowned scientists. The Eastern Hepatobiliary Surgery Hospital — an affiliated teaching hospital of the Second Military Medical University — Shanghai Jiao Tong University and Fudan University jointly produced a feasibility report for the National Center for Liver Cancer and the report was formally approved by the National Development and Reform Commission in 2010.

Construction of the National Center for Liver Cancer then commenced, backed by an investment of 400 million RMB. The centre covers an area of 30,450 square metres and includes a scientific research building, an academic exchange centre, experimental animal facilities and an administration



building. The infrastructure will be completed at the end of 2015. Wang will be the founding director of the centre, which is currently affiliated with and managed by the Eastern Hepatobiliary Surgery Hospital.

### Detection and prevention

The National Center for Liver Cancer will focus on seven key aspects of research: early detection and prevention of the transition from hepatitis to cirrhosis to liver cancer; screening of liver cancer biomarkers and related applications in early diagnosis, molecular typing and personalized treatment; study of the molecular mechanism of the development and progression of liver cancer, drug target identification and targeted treatment; research of the molecular mechanisms associated with recurrence and metastasis and of intervention methods for liver cancer; systematic regulation of the relationship between the host and microenvironment; exploration of new approaches for biotherapy of liver cancer; and novel comprehensive approaches for treating liver cancer.

To achieve its ambitious research goals, the centre will establish six core platforms and three large-scale shared facilities. The six core platforms will mainly consist of laboratories for diagnostic biomarker screening and molecular typing, pathogen-host interaction analysis, disease molecular network modulation, discovery and development of drug targets, biological therapy and evidence-based medicine. The large-scale shared facilities will include a biobank and a big-data information centre, a model animal centre and a polymerase-chain-reaction-certified experimental platform.

Wang and her group will focus on linking basic research findings with clinical investigations, with the aim of improving outcomes for hepatocellular carcinoma patients. The group has already made significant contributions to the understanding of molecular mechanisms of tumorigenesis, signalling network regulation and characterization of novel liver cancer biomarkers, which are directly related to the development of new therapies for patients. The group found that the protein Glypican-3 can be used as a



biomarker for diagnosing hepatocellular carcinoma, especially for alpha fetoprotein (AFP)-negative patients. This discovery will significantly improve the early detection of hepatocellular carcinoma.

The researchers have also established a tissue bank of liver cancer, which has more than 10,000 samples, and a systematic information system. They have published more than 160 papers in top journals in this field, including *Cancer Cell*, *Nature Cell Biology*, *Nature Communications*, *The Journal of Experimental Medicine*, *Gastroenterology*, and *Hepatology*. They have also received international and regional recognition, including the National Awards for Science and Technology Progress (Innovation Team Award), the National Natural Science Award (Grade II) and the Shanghai Natural Science Award (First Prize).

The centre also plans to take its research in other directions. These include studying the mechanism and prevention of malignant transition from inflammation to tumour and the facilitation of early diagnosis and intervention of this transition; regulation of the tumour microenvironment; metabolic regulation and malignant tumours; multicentre verification and clinical application of screening biomarkers; validation of drug targets based on research and development; identification and function research of cancer stem cells; as well as basic research, translational medicine and precision medicine based on targeting tumour heterogeneity.

### Invitation to talented researchers

Led by Wang, the National Center for Liver Cancer will build its strength in clinical research by taking advantage of clinical facilities in Shanghai, especially those at the Eastern Hepatobiliary Surgery Hospital, and by working closely with the basic liver cancer research teams at Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. The centre principally will strive to solve bottleneck issues in the prevention, diagnosis and treatment of cancers, especially liver diseases in China.

To achieve these long-term goals, the centre intends to recruit talented researchers from around the globe, including independent principal investigators and other scientific and technological personnel. It is eager to receive applications from talented researchers within China and overseas who are working in the field of liver cancer and oncology research. The National Center for Liver Cancer welcomes scientists at different levels, ranging from leading established scientists in the field to talented young researchers.

国家肝癌科学中心  
National Center for Liver Cancer

### Contact

**E-mail:** hywangk@vip.sina.com;  
**Tel:** +86-21-81875605  
**Fax:** +86-21-65566851  
**Address:** NO.700, North Moyu Road,  
Jiading District, Shanghai,  
201805, China.



# OSAKA UNIVERSITY

A comprehensive research university  
with world-class achievements

Cited papers

Chemistry **11<sup>th</sup>**

Immunology **16<sup>th</sup>**

Biology & Biochemistry **31<sup>st</sup>**

Material Science **33<sup>rd</sup>**

Physics **37<sup>th</sup>**

Molecular Biology & Genetics **39<sup>th</sup>**

These rankings are for universities only based on  
the Essential Science Indicators™ database.

---

---

**Japan's Most Innovative University**

**18<sup>th</sup> Most Innovative in the World**

By Reuters Top 100 Innovative Universities 2015

---

---

← **nature INDEX**

大阪大学  
OSAKA UNIVERSITY

[www.osaka-u.ac.jp](http://www.osaka-u.ac.jp)

Research at Osaka University

ResOU

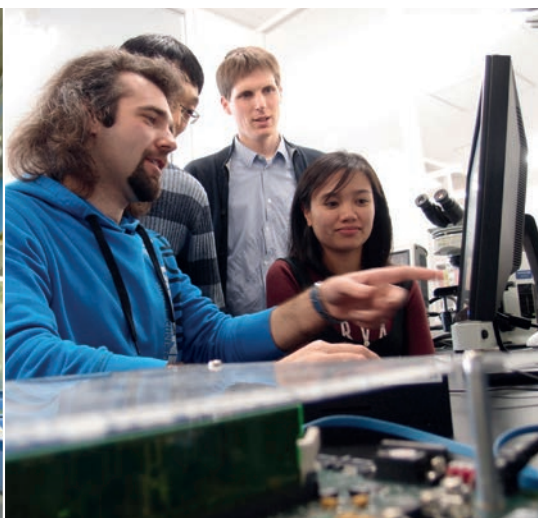
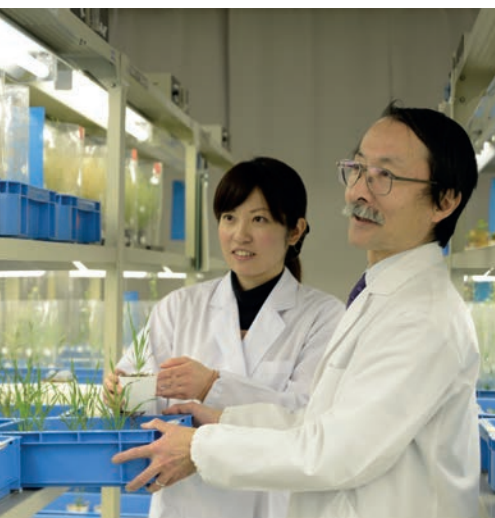
<http://resou.osaka-u.ac.jp>

Japan Nobel Prize winner,  
Hideki YUKAWA's favorite blackboard

Professor Hideki Yukawa, the first Japanese Nobel Laureate, served as a researcher and an assistant professor at Osaka University (formerly Osaka Imperial University). He first announced the existence of mesons in his thesis for which he attained his doctorate from Osaka University. This announcement eventually led to his receipt of the Nobel Prize in Physics in 1949.

The blackboard that Professor Yukawa regularly used in his laboratory while serving as a visiting professor at Columbia University in the U.S. was given to Osaka University's Graduate School of Science. This blackboard encourages our students to deepen their discussions and lead the way to the future of physics.





RIKEN

# Science for survival

**RIKEN's new president is asking its scientists to think about changing the world and they are rising to the challenge**

**Y**oshinori Tokura has a passion for skyrmions, which sounds a bit fantastical. A skyrmion is only about 20 nanometres in diameter, but it consists of thousands of spins that form a vortex in a material — “like a typhoon,” as Tokura puts it. “Once created, a skyrmion cannot be easily destroyed,” he says. “It can flow freely, acting as a high-energy particle.”

For Tokura, who directs the RIKEN Center for Emergent Matter Science (CEMS), skyrmions are a perfect example of an emergent entity, forming a topologically defined, independent object whose characteristics altogether surpass those of its component parts. They are a discovery of fundamental research and yet promise to lead to real-world applications — more efficient memory devices. They also represent the ideal that RIKEN is pursuing with success: the marriage of curiosity-driven basic research with potentially game-changing applications in areas of critical importance for our global future. CEMS and the RIKEN Center for Sustainable Resource Science (CSRS), both established in 2013, have been tasked with finding radical solutions to some of the world's most pressing

problems, including energy supply and environmental protection. That bar has been set even higher since April, when RIKEN entered a new era under the leadership of President Hiroshi Matsumoto.

## ‘Shaking things up’

According to Matsumoto, ‘sustainability’ is a bit of a misnomer for his vision of RIKEN's mission. As a concept, sustainability leaves people complacent, he says, offering the example of water, which people know is in short supply but do little to conserve. “The globe is almost broken,” he says. “We need to make people realize it's a matter of survival. Only then will people realize the significance of science.”

Matsumoto combines a deeply philosophical management style with an appreciation of social and economic realities and, above all, a respect for what science can achieve. “It's the only way we will survive, with the constantly increasing population in the developing world coming to expect a better quality of life,” he says. “RIKEN should offer something. We need to shake things up.”

He is shaking things up with a five-point plan that calls for RIKEN to become a global hub for innovation, where its researchers can form collaborations with those from universities and industry around the world.



RIKEN President Hiroshi Matsumoto.

He is encouraging more joint RIKEN appointments for university researchers and pressing researchers to find interdisciplinary opportunities to expand their research. “RIKEN will be the place where exchanges take place,” he says. His plan also calls for more foreign researchers — from the current 19 per cent to 30 per cent — and the creation of a fully bilingual RIKEN.

Perhaps most importantly for researchers, Matsumoto wants to ensure firm support for RIKEN scientists. He plans to increase the number of permanent staff and implement a tenure track system. He is seeking to increase researcher contracts from five to seven years; a period he says is required for many projects to bear fruit. He also plans to launch a similar programme



to one he introduced while president of Kyoto University in which a few select scientists will be given generous funding and set free to do research anywhere in the world.

With the benefits come expectations. "Excellency in science is not enough. Before being a scientist, you must be a human," he says. He tells his scientists to always consider how they can contribute to society: "RIKEN is a place where you can help change the world."

CEMS and CSRS are good examples of how this mission is being carried out.

### Carbon, nitrogen and metals

CSRS, which is directed by Kazuo Shinozaki, has researchers from various fields ranging from microbiology to catalytic chemistry, from plant science to chemical biology. Research at CSRS is focused on three targets, all of which require interdisciplinary collaboration and aim to create a society based on renewable resources and energy. Researchers engaged in the carbon initiative are enhancing photosynthesis in order to use carbon dioxide, which would otherwise contribute to global warming, to create useful materials. Research projects focused on nitrogen aim to find less fossil-fuel-intensive methods of creating ammonium and to create crops that produce high yields even with low nitrogen fertilizers. The metallic elements initiative aims to create cheap and readily available catalysts for chemical synthesis. It is also targeting a means to recover metals using moss or microorganisms. In addition, CSRS has a biomass engineering project geared towards translational research in collaboration with industry. To become more international, the centre has adopted English as its official language.

Experimental results are already capturing global attention. Zhaomin Hou's discovery that a novel titanium polyhydride complex can break the strong bond between two atoms of nitrogen ( $N_2$ ) will open a way to break the  $N_2$  bond at room temperature, a crucial step towards a long-sought method for producing ammonia more cheaply. That work followed closely on Hou's discovery that the titanium polyhydride complex can be used to break carbon-carbon bonds in benzene, opening up the possibility of previously

unimaginable applications for the whole range of aromatic molecules.

CSRS scientists are also making breakthroughs relating to food safety and security. Last year, Kazuki Saito reported the ability to silence a potato gene involved in the production of a toxin. Ryoung Shin is building on the discovery of a chemical that binds to caesium to make plants that will not accumulate the radioactive substance, which will contribute to Fukushima remediation efforts. Shinozaki dreams of understanding complex gene networks well enough to produce drought-resistant plants. "This will be very important for the future of food security," he says.

As director, Shinozaki strikes a balance by encouraging CSRS researchers to think broadly about the scope of their science and its applications. "We give them freedom to pursue research in a particularly interdisciplinary way," he says. "We especially encourage young researchers to take on this challenge."

### Ubiquitous skyrmions

At CEMS, Tokura is already eyeing applications for skyrmions. Skyrmions were once just a theoretical hypothesis and were then thought to occur only under special circumstances, but Tokura used cutting-edge Lorentz electron microscopy to show that skyrmions are ubiquitous in conventional magnets. This discovery opened the door to using skyrmions as information carriers in next-generation magnetic memory devices.

"People are excited; they're jumping into the field, especially engineers in spintronics," says Tokura. The skyrmion research moved ahead only because of close collaboration with other groups at the centre, bringing together theoreticians and those who could design the structures of new materials, those who could make the materials, and those who could observe skyrmion properties. "There are no barriers between laboratories within the centre," he says.

Another representative example of interdisciplinary emergent matter research at CEMS is the development of aqua materials, or 'watery plastics', made from 95 per cent water. By linking water molecules together with tiny amounts of clays and organic macromolecules, CEMS researchers have

## Fast facts:

- **RIKEN, a national R&D institution founded in 1917, is Japan's largest and most comprehensive research organization for basic and applied science.**
- **RIKEN conducts world-leading research in developmental biology, neuroscience, drug development, plant science, nuclear physics, solid-state physics, photonics, chemistry, materials science and supercomputing.**

created aqua materials that are environmentally benign and moldable with special biological and optical functions.

With close to 30 per cent foreign staff and a host of international and industrial collaborations, CEMS is already starting to make headway on its formal mission of contributing to a sustainable society through big innovation.

Aside from his own goals with skyrmions, Tokura has laid out "Big Four" goals that require massive innovation: tripling the efficiencies of superconductors, solar cells, batteries and thermoelectric materials used in refrigeration. These goals are ambitious — "a bit crazy," says Tokura — but they could radically rewrite the economics of energy. "We always have to keep these kinds of issues in mind," says Tokura. "Then, as long as we are allowed to do very basic research, we have a chance to make a big leap."



## Contact

[www.riken.jp/en](http://www.riken.jp/en)

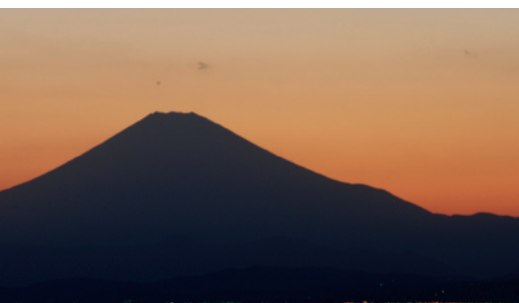
[www.riken.jp/en/careers](http://www.riken.jp/en/careers)

[www.facebook.com/RIKEN.english](https://www.facebook.com/RIKEN.english)

[www.twitter.com/riken\\_en](https://www.twitter.com/riken_en)

Tel: +81 48 462 1225

E-mail: [pr@riken.jp](mailto:pr@riken.jp)



**SOKENDAI (The Graduate University for Advanced Studies)**

# Incubating excellent researchers at Japan's finest institutes

In conjunction with its numerous 'parent institutes', SOKENDAI offers comprehensive doctoral programmes that provide students with the opportunity to study in an ideal academic environment while enjoying access to cutting-edge facilities and being steered by the finest researchers.

Established in 1988 as Japan's first independent graduate university, SOKENDAI offers three- and five-year doctoral programmes. SOKENDAI is unique in that it

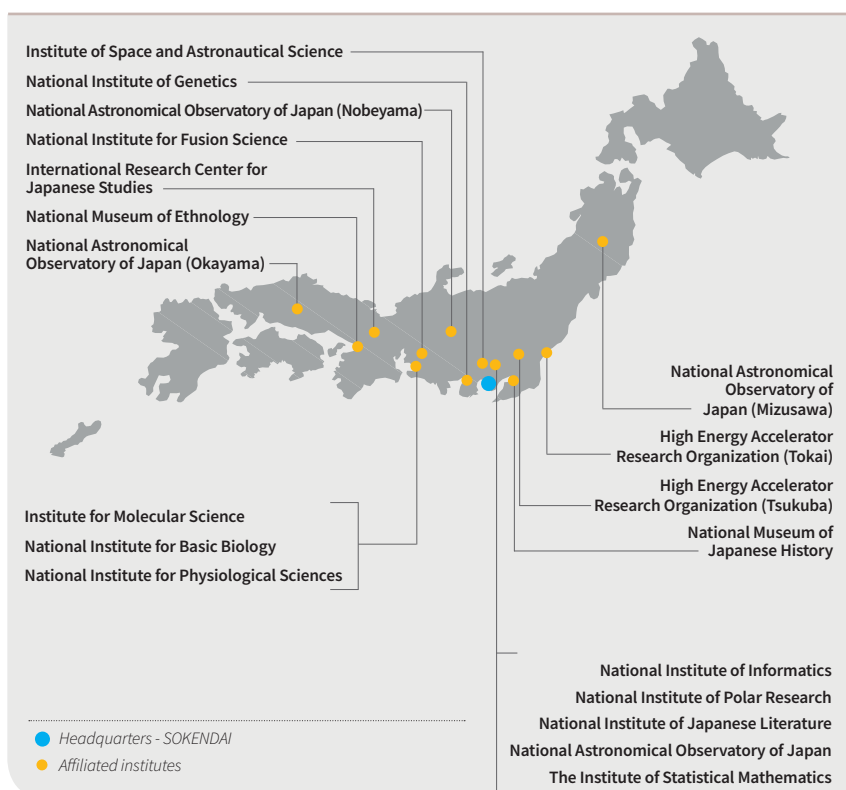
is affiliated with world-renowned, international research bases, known as Inter-University Research Institutes (IURIs). Most professors and students at SOKENDAI conduct research and study as members of IURIs, which provide ideal environments for these activities. In particular, students have access to the large-scale experimental and analytical facilities of the IURIs that host their schools.

This system allows a favourable staff-to-student ratio of 2.5 faculty members

per student to be achieved, so that it offers a rich educational environment for doctoral studies. In addition, being an independent graduate university, SOKENDAI attracts students from various universities, backgrounds and nationalities.

Graduate education at SOKENDAI covers a wide range of fields including cultural studies and the natural sciences. Teaching is conducted at the 20 departments of the 18 IURIs, where joint, cross-field research is actively pursued in conjunction with other institutes in Japan and overseas. This environment provides the optimum conditions for students, allowing them to acquire high levels of expertise, a cross-disciplinary perspective and an international outlook. This fosters excellent researchers with outstanding knowledge and a broad perspective.

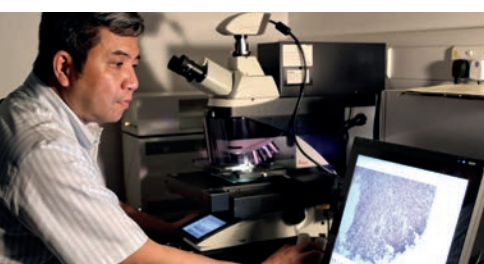
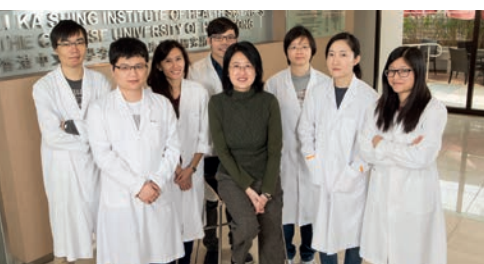
SOKENDAI aims to establish a new research graduate school that accepts both bachelor and master graduates from diverse academic backgrounds who are ready to take on a challenge.



 National University  
**SOKENDAI**  
 The Graduate University for Advanced Studies

## Contact

**E-mail:** [kouhou1@ml.soken.ac.jp](mailto:kouhou1@ml.soken.ac.jp)  
**Tel. no:** +81-46-858-1500  
**Fax. no:** +81-46-858-1542  
**Website:** [www.soken.ac.jp](http://www.soken.ac.jp)



## Faculty of Medicine, Chinese University of Hong Kong

# Turning scientific advances into innovative clinical practices

**Established in 1981, the Faculty of Medicine at the Chinese University of Hong Kong (CUHK) now has 19 schools and departments, with over 250 academic staff members. The faculty's expertise in cutting-edge translational cancer and genomic research is essential to its mission to improve human health through scientific advances.**

The Faculty of Medicine, CUHK has established multiple research centres and institutes with a special focus on translational cancer and genomic research, three examples include:

- **Li Ka Shing Institute of Health Sciences (LiHS):** A key translational medicine research institute in Hong Kong, the LiHS opened in 2007 and currently has over 300 researchers. Its key focus area is translational genomics, particularly in the area of non-invasive prenatal testing and cancer.
- **Sir Y. K. Pao Cancer Centre:** Opened in 1994, the centre is the first dedicated facility in Hong Kong for cancer research and clinical services. It houses the State Key Laboratory in Oncology in South China, which focuses on cancers common in Hong Kong, such as liver, lung and nasopharyngeal cancers.

Both the LiHS and the Sir Y. K. Pao Cancer Centre are located in the Prince of Wales Hospital, the main teaching hospital of the Faculty of Medicine, CUHK.

- **School of Biomedical Sciences (SBS):** The SBS was created in 2009 through the unification of the Departments of Anatomy, Biochemistry (Medicine), Pharmacology and Physiology. The 42 professorial staff members at the SBS conduct research that falls under five thematic research programmes: (1) cancer and inflammation; (2) neuro-degeneration, -development and

repair; (3) reproduction, development and endocrinology; (4) stem cell and regeneration; and (5) vascular and metabolic biology. With its recent relocation to a new state-of-the-art facility in 2012, the school is reaching new heights of excellence in biomedical research.

The Faculty of Medicine at the CUHK has established numerous international ties, winning regional and international recognition for its work in translational genomic research. Highlights of this research include:

### **Circulating cell-free nucleic acid diagnostics**

The detection of circulating nucleic acids — DNA and RNA molecules released from the cells of multiple body tissues into the blood — provides an opportunity to develop non-invasive blood-based diagnostics for a variety of conditions. The circulating nucleic acids laboratory at the CUHK, led by Dennis Lo and Rossa Chiu, is dedicated to the development of novel molecular diagnostics for the assessment and monitoring of human diseases and pathologies.

Lo and Chiu's team contributed to the development of non-invasive DNA-based approaches for the prenatal diagnosis of fetal chromosomal and genetic diseases by demonstrating the presence of fetal DNA in the circulation of pregnant women. Together with King's College Hospital in London and the VU University Medical Centre in the Netherlands, the CUHK group pioneered an approach based on massively parallel sequencing of cell-free DNA molecules in maternal blood that is effective for the detection of fetal Down syndrome. Soon after, the test became available for clinical use and is now offered in over 100 countries. The CUHK



team was also the first to achieve the direct determination of the fetal genome through the analysis of maternal blood, which has paved the way for non-invasive diagnosis of fetal single-gene diseases. The team has collaborated extensively with Charles Cantor at Sequenom to bring this technology to the clinic.

The CUHK team was also the first to show that genetic and genomic changes of solid cancers can be detected in a genome-wide fashion by massively parallel sequencing analysis of cell-free DNA in blood. This non-invasive molecular scan is able to detect genomic changes associated with a range of malignancies, including cancers of the liver, lung, colon, breast and nasopharynx. The team's work has been widely published in high-impact journals and has been awarded multiple international honours, including election to the Royal Society and the US National Academy of Sciences.

### Liver Cancer Genome Project

Liver cancer is a highly aggressive cancer that is prevalent in China and Southeast Asia. In 2011, Nathalie Wong and her research team received a HK\$45 million grant from the Hong Kong Research Grants Council to undertake a 'Liver Cancer Genome Project'. The project is a collaboration between clinicians and basic researchers from the CUHK, the University of Hong Kong, the Hong Kong University of Science and Technology and the Beijing Genomics Institute in Shenzhen that aims to comprehensively delineate whole genome and whole-transcriptome somatic variations in liver cancer.

The project focuses on large-scale genome-wide analyses to define genetic events that distinguish tumours from cirrhosis and progression to metastatic disease. Using next-generation sequencing, Wong's team recently illustrated the mutagenic effects of hepatitis B virus (HBV) integration in the development of liver cancer. Through transcriptome sequencing, the team also demonstrated the presence of viral-human chimeric transcripts from the site of genome integrations and helped to elucidate the oncogenic effect of viral insertion. The project has thus provided a foundation for future research into effective control strategies for the disease.

### Nasopharyngeal cancer research

Nasopharyngeal carcinoma (NPC), a distinct type of head and neck cancer, is a major healthcare problem in Hong Kong and South China. For the past two decades, a team led by Kwok-Wai Lo, an international leader in NPC genome research, has sought to decipher the molecular basis of NPC and identify novel targets for this deadly disease. The group collaborates with the Dana-Faber Cancer Institute in the US and the ILCHUN Genomic Medicine Institute in South Korea, and works closely with other internationally renowned research groups. Their pioneering work has provided significant insight into the molecular biology of NPC development.

Recently, Lo and his team discovered a novel recurrent transforming fusion gene in a subset of patients — the first evidence to support the role of gene rearrangement in the genesis of NPC. To offer a molecular basis for the development of



promising disease control strategies for NPC, Lo's group has initiated a new project to uncover vital genetic changes through large-scale whole genome sequencing and bioinformatic analysis. The team plans to systematically define the driver mutations, identify key molecular targets and elucidate underlying biological mechanisms to pinpoint novel biomarkers and therapeutic targets. The project is expected to contribute to the development of personalized treatment strategies for NPC patients.

### Clinical cancer research

The State Key Laboratory of Oncology in South China is built on the CUHK's long-established partnership with Sun Yat-sen University, Guangzhou, in cancer research, as well as its long tradition of cancer research and care in the local community. Led by Anthony TC Chan and Tony SK Mok, the laboratory has established current standards of care in the management of locoregionally advanced NPC and lung cancer.

Mok's team focuses on biomarker and molecular targeted therapy in lung cancer. With the discovery of epidermal growth factor receptor mutation, Mok designed and led the IPASS (Iressa Pan-Asia Study), which confirmed the role of this mutation as a predictive biomarker and demonstrated the superiority of targeted monotherapy over doublet chemotherapy. Mok has also led a number of other multinational clinical trials that have contributed to the current standard of practice for the management of advanced-stage lung cancer.

Focusing on locoregionally advanced NPC, Chan and his team reported the first prospective randomized study in one





endemic area to establish concurrent cis-platin-radiotherapy as the current standard of care. Collaborating with Cancer Research UK, they developed a therapeutic Epstein-Barr virus (EBV) vaccine to boost *in vivo* T cell responses to EBV proteins expressed by tumour cells. The pilot study established the safety and immunogenicity of the EBV vaccine. A multinational phase 2 efficacy study is ongoing.

### Digestive diseases and gastrointestinal cancer studies

The Institute of Digestive Disease (IDD), established in 2006 by Joseph JY Sung, seeks to improve the diagnosis and treatment of digestive diseases through basic, translational and clinical research. Comprising a multidisciplinary group of investigators from a wide range of internationally recognized academic fields, IDD's ultimate goal is to improve patient survival through the early detection and treatment of gastrointestinal cancers.

The IDD has made a number of important discoveries related to the treatment of digestive diseases, including a technique for the management of peptic ulcer bleeding, led by Francis KL Chan, and insight into the treatment of HBV infection, led by Henry LY Chan. In addition, the research laboratory led by Jun Yu has pioneered the use of microRNAs as novel non-invasive biomarkers for colorectal cancer, and has discovered novel tumour suppressor genes in gastric cancer and new mutation signatures in colorectal and gastric cancers. The IDD's research has been published in a number of international peer-reviewed journals, including the *New England Journal of Medicine* and *The Lancet*. Of IDD's over

## Discovered at the CUHK

- **Dennis Lo and Rossa Chiu noninvasively determined the fetal genome, DNA methylome and transcriptome through maternal blood analysis.**
- **Nathalie Wong identified a novel mechanism of viral-host fusion sequence HBx-LINE1 in liver oncogenesis.**
- **KW Lo generated a high-resolution genome map of EBV-associated nasopharyngeal cancer.**
- **Anthony Chan and Tony Mok led international multi-centre anti-PDL1 immunotherapeutic studies in nasopharyngeal and lung cancers.**
- **Jun Yu discovered novel molecular events and biomarkers of gastrointestinal cancer through genome sequencing and bio-functional analyses.**
- **HC Chan discovered a novel role of cystic fibrosis transmembrane conductance regulator in insulin secretion and cystic fibrosis-related diabetes.**

1,600 publications, 300 were published in high-impact journals.

The IDD has been honoured with 42 prestigious national and international awards. Core IDD team members serve on editorial boards of major medical journals and advisory boards of major industrial partners and have secured 120 external local, national and international research grants of over HK\$350 million in the past 5 years.

In addition, The State Key Laboratory of Digestive Disease was established at the IDD in 2013, in partnership with the Fourth Military Medical University in mainland China. The IDD's strong international network, enables collaborations with key leaders in the field around the world.

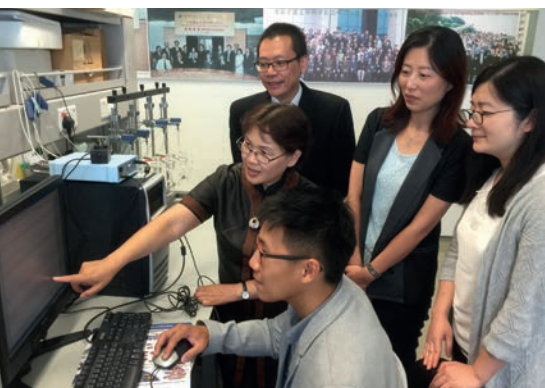
### Epithelial cell biology research

Epithelial cells form a diverse group of tissues that cover or line almost all body surfaces, cavities and tubes, such as the lungs, liver, kidneys and reproductive tract. Defects in epithelial cell function are therefore associated with a wide spectrum of disorders and diseases, including hypertension, diabetes, infertility and cancer. Recognizing the importance of epithelial cells, researchers at the CUHK and the Academy of Military Medical Sciences joined forces in 1999 to establish the first and leading Epithelial Cell

Biology Research Centre (ECBRC) in China, with help and support from the National Natural Science Foundation of China.

Since its establishment, the ECBRC has built a multidisciplinary technology platform and has stayed at the forefront of epithelial cell biology research. This cross-disciplinary approach, together with a network of over 50 collaborating laboratories worldwide, have led to ECBRC's high-impact publications in leading journals, such as *Nature Medicine*. Among the ECBRC's many notable discoveries, ECBRC researchers have identified the role of epithelial ion channels in reproduction, mechano-sensing, cancer development and insulin secretion, and have illustrated the importance of defensins in sperm function and male fertility.

With a strong team led by Hsiao Chang Chan, the ECBRC is engaged in a wide variety of basic and translational research. From investigation of the ion channels in signal transduction to researching risk factors of assisted reproduction technology, the ECBRC's research has numerous potential applications in the diagnosis and treatment of infertility and other diseases. The ECBRC embraces the team spirit and is open to further collaborations, including joint ventures with the industrial sector to promote translational research. "Together, great things are possible," Chan says.

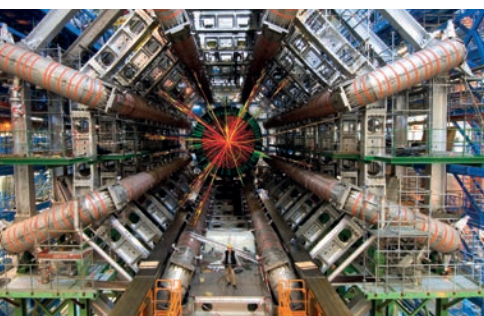
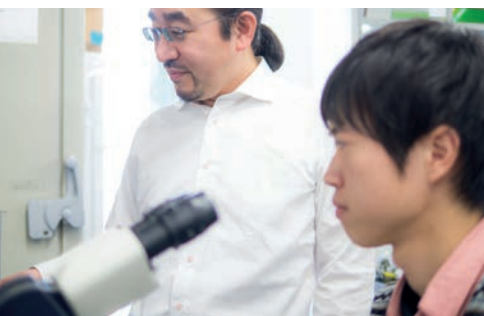
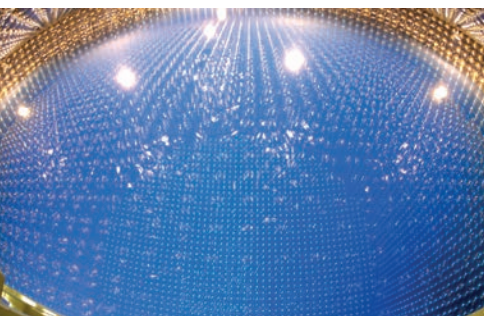


## Contact

**Email:** [emmylee@cuhk.edu.hk](mailto:emmylee@cuhk.edu.hk)  
**Tel:** 852-2632-3799  
**Fax:** 852-2649-0208  
**Web:** [www.med.cuhk.edu.hk/eng](http://www.med.cuhk.edu.hk/eng)



香港中文大學醫學院  
**Faculty of Medicine**  
 The Chinese University of Hong Kong



## The University of Tokyo

# Collaboration: the key to competitiveness

**The University of Tokyo's new president has a global vision that builds on a long history of accomplishment by inspired researchers. The joint awarding of the 2015 Nobel Prize for Physics to Takaaki Kajita, Director of the university's Institute for Cosmic Ray Research, is just the latest manifestation of the university's enthusiasm for science.**

**M**akoto Gonokami, who took over as president of the University of Tokyo in April, has a new vision for Japan's premier university.

The university is a place where the creativity of Japan melds with global trends. Steady economic growth and peace have provided a supportive backdrop for outstanding research results that are highly regarded all around the world. Kajita's Nobel Prize for the discovery of neutrino oscillations shows just how profitable support from the university can be.

Gonokami wants to reinforce and intensify the university's mission. He wants to encourage researchers to pursue original ideas and provide the flexibility and the research environment needed to do so. He wants the university to be a place where the boundaries between nations, cultures, generations and academic disciplines are lowered. Only then, he believes, will academic researchers acquire the broad vision needed to tackle issues such as the depletion of natural resources, environmental damage, energy security, infectious disease and Japan's declining birthrate.

Gonokami plans to make the university a "global base for knowledge

collaboration" that cultivates a new class of "knowledge professionals." One of his key initiatives is a Graduate School of International Excellence, which will blend the sciences and humanities. The graduate school will create a flexible environment in which to pursue interdisciplinary careers.

"I would particularly like students who have their hearts set on learning and young researchers to experience the excitement and joy of knowledge," says Gonokami.

Gonokami also plans to enhance internationalization by creating a more flexible system to bring in more visiting lecturers, researchers and students from around the globe. This will build on ongoing internationalization efforts.

To lower barriers between academia, industry and government with a view to capitalize on the fruits of scientific research, Gonokami is promoting collaborative research with industry that benefits society, the establishment of venture companies and a reorganization of the university's intellectual property management system.

One example of such barrier-breaking collaborative initiatives is PeptiDream, formed in 2006 by Hiroaki Suga as an outlet for his pioneering peptide research. PeptiDream's success grew from an artificial ribozyme that can produce peptides with the potential to act as drugs. The system can even introduce amino acids that are not present in naturally occurring proteins to the peptides. This ability enables peptides with better pharmacological and chemical properties to be created. Crucially, the system is fast, efficient and cheap, pumping out trillions of peptides a week

Photo 5: ©2014 CERN

Photo 4: ©Asahi Shimbun Company

Photo 2: ©Kamioka Observatory, ICRR, University of Tokyo

and providing drug companies with hundreds of potential targets.

The technology, which took twelve years to develop, required another four years to be accepted by industry. “At the beginning, no one thought macrocyclic peptides could be drug candidates,” Suga says. Now, PeptiDream counts most major multinational pharmaceutical companies on its list of partners, including AstraZeneca, Novartis, GSK and Lilly. And despite having a mere 35 staff, it is valued at US\$1.5 billion.

Support from the university was crucial. It provided space in an incubation centre and the university technology-transfer office provided all the legal support needed, enabling Suga to maintain a cutting-edge laboratory. “Visitors from other countries say, ‘Wow, you can do anything you want here,’” he says. An influx of talented researchers from overseas and within Japan provided the necessary interdisciplinary talent.

The key, however, is the original idea. “This is a unique technology — no one else can make that kind of molecule,” Suga says. “Once people see it, they are convinced.”

Suga dreams that one of his peptides will become a blockbuster drug. “I want to help patients. I want to meet unmet medical needs,” he says. PeptiDream’s peptides are in 20 drug pipelines, a figure that makes most pharmaceutical or biotechnology companies envious. “There are infinite possibilities to develop a new class of drugs. Non-druggable targets can become druggable,” he says.

Creating the company allowed Suga to straddle the border between academia and industry, remaining “relaxed and open-minded” enough to follow his curiosity towards its next discovery.

Crossing another kind of boundary, Shoji Asai’s research on the Higgs boson showed how well The University of Tokyo researchers, through collaboration and competition, are playing critical roles in the biggest scientific findings of the day. Asai’s team found three of the decay channels of the Higgs boson — channels through which the fleeting elementary particle was identified. The discovery was the “most exciting

moment of my career,” says Asai, “and one I will treasure.”

Asai’s team was competing with groups from other countries before eventually producing the most reliable and sensitive method for identifying the Higgs boson. “We had to fight hard to establish these methods,” he says.

At the same time, the team depended on international collaboration. The ATLAS detector used in the experiments has 150 million sensors and required the work of 3,000 scientists from around the world to create and operate it. “Without this collaboration, there would not have been a Higgs boson discovery,” Asai says.

The international project opened a window into the elementary particle as well as a window to the world for many scientists. More foreign researchers and graduate students have been coming to Japan, while Japanese with doctorates from the University of Tokyo “are spreading all over the world,” says Asai.

Aside from the generous funding and facilities, the university’s willingness to let him work freely has been essential, says Asai. Now, he will follow his next goal and ultimate dream — to build on the Higgs boson work to discover supersymmetry, something that could explain elusive dark matter.

The University of Tokyo has also been leading efforts, with the support of the government, to reform the research system in Japan so that more such discoveries are possible.

For example, a government-appointed international evaluation committee found that the university’s Kavli Institute for the Physics and Mathematics of the Universe (IPMU) has “led the way toward organizational reform of Japanese education” with its establishment of full faculty positions, a merit-based salary system and other management approaches rarely found in Japan. The committee recommended renewing the Kavli-IPMU’s five-year term noting that it had exceeded expectations. The centre can pursue research into the state of the Universe a trillion years into the future using Japan’s Hawaii-based Subaru telescope and investigate the tiny Universe prior to the Big Bang using the



Professor Makoto Gonokami, the president of the University of Tokyo

LiteBIRD satellite. The Kavli-IPMU’s success has proved that coming to Japan can be career boosting, says Institute Director Hitoshi Murayama.

In addition, the Ministry of Education, Culture, Sports, Science and Technology in Japan selected the university’s proposal of a Top Global University Project. The project will create collaborations that extend from research to education with strategically selected universities overseas. By cultivating “mobility, excellence and diversity,” the goal is to produce intellectual leaders with global outlooks. At the same time, the University of Tokyo will continue overhauling its own education system to become a role model for other Japanese universities. “I want students to use their experience as a source of intellectual nourishment that reinforces the willingness to develop themselves,” says Gonokami.



THE UNIVERSITY OF TOKYO

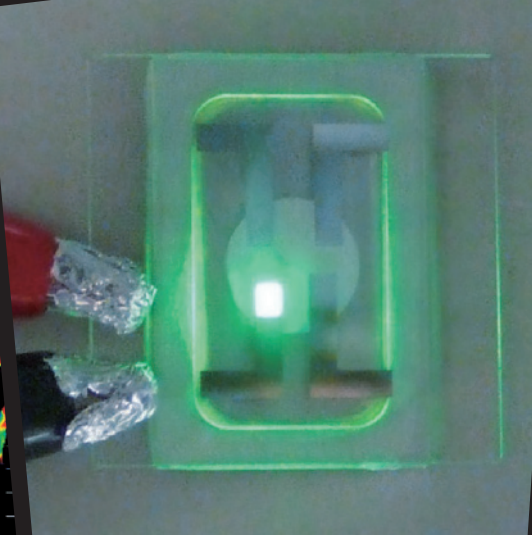
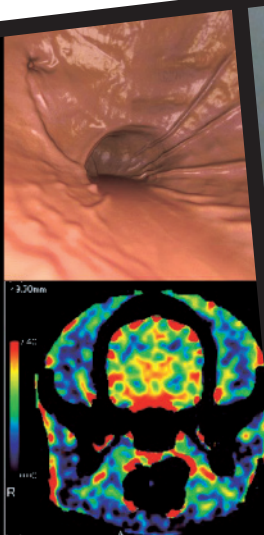
## Contact

**Website:** [www.u-tokyo.ac.jp/en](http://www.u-tokyo.ac.jp/en)  
[www.facebook.com/UTokyo.News.en](https://www.facebook.com/UTokyo.News.en)  
[www.twitter.com/UTokyo\\_News\\_en](https://www.twitter.com/UTokyo_News_en)  
[www.youtube.com/user/UtokyoPR](https://www.youtube.com/user/UtokyoPR)  
**Phone:** +81-3-3811-3393



# To fulfill our role as a University that brings Japan to the world

- ◆ to strengthen the advanced research capabilities that enable globally competitiveness
- ◆ to promote educational research that encourages interaction with the international community
- ◆ to be a driving force in bringing Japan's industry to the international community
- ◆ to develop sophisticated and innovative leaders

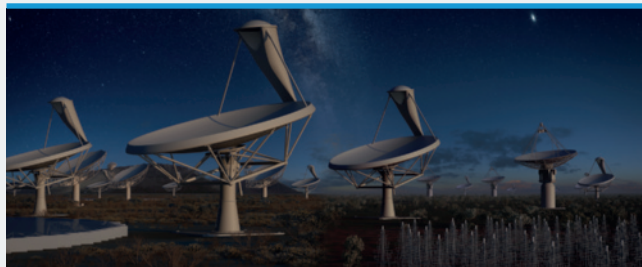






**The University of Cape Town's geographical vantage point at the tip of Africa offers an exhilarating research environment that combines excellence with impact.** As a leading research university (1st in Africa THE, QS, ARWU 2015; 120 in the world THE 2015), UCT is one of the best places in the world to research Africa-specific problems, from the chemistry of malarial drug discovery to the development of urban Africa. We have strong collaborative networks across the globe, and often form a nexus of partnerships between researchers in the global north and south, particularly Africa. It is therefore not surprising that we attract researchers — from postgraduates to leaders in their field — from all over the continent and the world.

SKA, WIKIMEDIA COMMONS

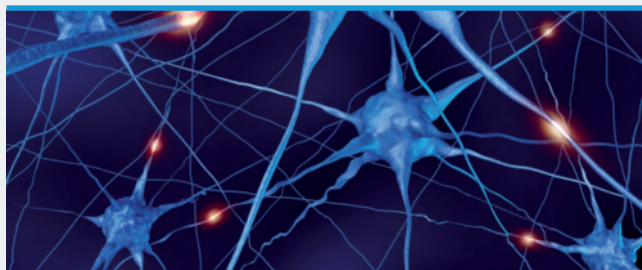


#### Experts in the Southern Skies

UCT is fast becoming Africa's astronomy and cosmology hub. It includes global leaders, such as George Ellis, who co-wrote *The Large Scale Structure of Space-Time* with Stephen Hawking; Michael Feast, who published his first paper in *Nature* at 21 and his most recent at the age of 87; and Russ Taylor, who is leading our involvement in the SKA project and the big data challenges it brings.

#### Conserving biodiversity

A recent paper in *Nature Plants* describes a local plant that engages in what is possibly the best example of faecal mimicry for seed dispersal anywhere in the world. This is typical of the work we do in our hotspot of biodiversity, with extensive interdisciplinary expertise in conservation, climate adaptation and community conflict.



#### Big in brains

Some neurological disorders, such as neuro-HIV, have far greater prevalence in Africa than elsewhere. UCT draws together cutting-edge technology and skills from a range of neuroscience disciplines to create a web of expertise that integrates laboratory, clinic and community, and opens new ways for Africa to contribute to the global body of knowledge in this rapidly advancing field.





# “Surrey pursues research excellence today for the world tomorrow”



## Truly International

- » An international staff and student body from over 145 countries
- » Achieved 39th position in the Times Higher Education list of the top 100 most international universities in the world 2013-14 and 247th in the 2016 QS world university rankings.
- » A strong student and staff mobility programme with over 85 international higher education partners and more than 2,300 placement providers
- » Strong academic connections with BRIC countries with 1700 students at our campus in Dalian, China in partnership with Dongbei University of Finance and Economics
- » Founding member of the University Global Partnership Network with University of São Paulo and North Carolina State University

## World Leading Teaching and Research

- » A comprehensive University with a global research footprint – 55% of our publications involve international collaborators
- » A new School of Veterinary Medicine launched in 2014
- » Ground breaking research in water, satellites and ion-beam technology recognised through three Queen's Anniversary Prizes
- » High profile research projects in sustainable development and renewable energy
- » creating the first 5G test bed in the world in 2015
- » Long term relationship with the Organisation for the Prohibition of Chemical Weapons – winners of the Nobel Peace Prize 2013

For more information, visit:  
**[surrey.ac.uk/nature](http://surrey.ac.uk/nature)**







# Zhejiang University

## SEEKING THE TRUTH AND PIONEERING NEW TRAILS

Zhejiang University (ZJU), one of the leading research-intensive universities in China, is located in the beautiful city of Hangzhou. Research at ZJU spans 12 academic disciplines: philosophy, economics, law, education, literature, history, art, science, engineering, agriculture, medicine and management. Of the 22 subjects listed in Thomson Reuters Essential Science Indicators, ZJU is rated in the top 1% in the world for 17.

As a national university, ZJU is building itself into an innovative and comprehensive research institution of world-class standing. ZJU seeks the truth and pioneers new trails. It is dedicated to creating and disseminating knowledge, advancing civilization, serving and leading society, and promoting national prosperity, social development and human progress.

ZJU sincerely invites talents to join it so that together we can create a glorious future.

### Contact:

Talents Office, Human Resources Department  
Zijingang Campus, Zhejiang University  
No. 866, Yuhangtang Road, Hangzhou  
Zhejiang, 310058, P. R. China

**Tel:** +86-571-88981390, 88981345, 88981656,  
88981453 (Postdoc)

**Fax:** +86-571-88981976

**E-mail:** tr@zju.edu.cn

<http://talent.zju.edu.cn>

## Why did you choose Zhejiang University?



"Zhejiang University acts as a bridge to the exciting and fast-growing Chinese market. It allows researchers to harvest real-life problems and provides abundant

opportunities for applying research outcomes to real systems. Besides providing opportunities to work with the best students in China, Zhejiang also has an excellent research environment that nurtures world-class research."

**Professor Wen Yuan Xu, College of Electrical Engineering**

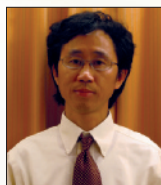


"For a young researcher, China is a fascinating country to begin an independent career. The creativity and pace of modernization will impress everyone who visits,

and having a position here gives you the opportunity to focus on what really matters: research and the education of a new generation of scientists. Zhejiang University is one of the top universities in China and provides starting faculty with excellent working conditions.

The Department of Chemistry supports its young professors wherever it can, gives them much freedom and welcomes interdisciplinary collaboration. I am happy to have brilliant and helpful colleagues as well as smart, hard-working students."

**Research Professor Simon Duttwyler, Department of Chemistry**



"I decided to return to China in 2009 after working abroad for 15 years as I believe that China is the best place to find sustainable devel-

opment solutions to address our planet's limited resources. Indeed, the Yangtze River Delta, where Zhejiang University is located, is an area that has prospered continuously for thousands of years without major environmental decay. In particular, I enjoy the freedom to pursue my area of interest as a researcher and the open-minded atmosphere at Zhejiang University."

**Professor Xiaogang Peng, Department of Chemistry**



"As a former PhD student at Zhejiang University, I have always valued its strong support network, so after doing two postdoc programmes

here, when the opportunity came for me to stay and teach, I took it without hesitation. I am proud to be on the faculty of the College of Agriculture and Biotechnology and I feel like I can never repay my alma mater for transforming me from a young boy into the quality teacher that I am today."

**Lecturer Imran Haider Shamsi (Ying Lan), Department of Agronomy**



"Zhejiang University's location in the economically dynamic region of the Yangtze River Delta provided much opportunity for translational research and

a large patient base for biomedical research."

**Professor Lanjuan Li, Director of the State Key Laboratory for Infectious Diseases Diagnosis and Treatment**



"Although I officially joined Zhejiang University in 2011, I have been collaborating with researchers at Zhejiang University since 2002. I have found it to be a place with

great academic freedom, and I believe that's the reason why Zhejiang University is one of the best universities in China. In addition, I am grateful to be able to work in the beautiful city of Hangzhou, where I grew up. Zhejiang University is also strong in engineering, which provides a fantastic platform for physical science research and interdisciplinary collaboration — activities that are of particular importance to me."

**Professor Fuchun Zhang, Department of Physics**



**浙江大学**  
ZHEJIANG UNIVERSITY